DOCUMENT RESUME

ED 177 999 HE 011 907

AUTHOR Breland, Hunter M.
TITLE Population Validity and College Entrance Measures. Research Monograph Number 8.
INSTITUTION College Entrance Examination Board, New York, N.Y.
PUB DATE 79
NOTE 101p.
AVAILABLE FROM College Board Publication Orders, Box 2815, Princeton, NJ 08541

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Admission (School); *Admission Criteria; *College Bound Students; Comparative Statistics; Correlation; *Grades (Scholastic); Higher Education; Measurement; Minority Groups; Predictive Ability (Testing); *Predictive Validity; Predictor Variables; *Standardized Tests; Statistical Analysis; Tables (Data)

ABSTRACT

Published reports and papers pertaining to the validity of measures commonly used at college entrance are reviewed and summarized in this report. These measures consist of various types of academic quantitative indices used to describe the high school record and various types of academic tests. The data are categorized by sex and ethnic classifications where possible, and comparisons are made. The validity of inferences made about different populations from the high school record and test scores is considered. Investigated was the degree to which performance of populations were over- or under-predicted when the data used in making inferences were not limited to data for that population. The data indicate that both the high school record and standardized academic tests appear to be useful at college entrance. Data also indicate that the college performance of black populations, both male and female, has been consistently over-predicted by the traditional academic indices when predictions are based on data from white or predominantly white samples. Women were shown to be consistently under-predicted by the traditional measures when predictions were based on data from male or predominantly male samples. The data are extensively tabulated and correlation tables of predictor validity are appended. (Author/SF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Population Validity and College Entrance Measures

ED177999

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

CEEB

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 011 907

Also in this series:

The Validation of a Biographical Inventory as a Predictor of College Success. Research Monograph No. 1. Anne Anastasi, Martin J. Meade, and Alexander A. Schneiders (20080) 82 pages, $2.50

Organizational Measurement and Its Bearing on the Study of College Environment. Research Monograph No. 2. Allen H. Barton (254730) 82 pages, $2.50

Career Development: Choice and Adjustment. Research Monograph No. 3. David V. Tiedeman and Robert P. O'Hara (213810) 108 pages, $5

Career and Development: Self-Concept Theory. Research Mongraph No. 4. Donald E. Super, Reuben Starishevsky, Norman Matlin, and Jean Pierre Jordaan (213815) 95 pages, $4.50

Personality Measures in Admissions: Antecedent and Personality Factors as Predictors of College Success. Research Monograph No. 5. Morris I. Stein (270500) 69 pages, $2.50

The Measurement of Writing Ability. Research Monograph No. 6. Fred I. Godshalk, Frances Swineford, and William E. Coffman (251700) 84 pages, $2.50

Using Self-Reports to Predict Student Performance. Research Monograph No. 7. Leonard L. Baird (25701) 90 pages, $5

Orders accompanied by either a check made payable to the College Board or an institutional purchase order may be sent to:

College Board Publication Orders
Box 2815, Princeton, New Jersey 08541

Population Validity and College Entrance Measures

Research Monograph Number 8

by Hunter M. Breland

Research Psychologist
Educational Testing Service

The College Board, New York, 1979

Editorial inquiries concerning this book should be addressed to: Editorial Office, The College Board, 888 Seventh Avenue, New York, New York 10019.

Copies of this book may be ordered from College Board Publication Orders, Box 2815, Princeton, New Jersey 08541. The price is $7.50.

Cover photograph by Glenn Foss

Copyright © 1979 by College Entrance Examination Board, New York. All rights reserved.

Library of Congress Catalog Number 79-51738

Printed in the United States of America

Contents

Preface vii

Acknowledgments. viii

Abstract ix

Introduction xi
- Accepted notions of validity. xi
- Population validity xii

Problems in empirical validation. 1
- Criterion problems 1
- Restriction of range 2
- Attenuation resulting from unreliability 3
- Problems of pooled samples 4
- Over- and underprediction 5
- Problems in internal analyses 6
- Summary of problems in empirical validation 6

Regression comparisons 7
- Summary of regression studies 7
- Studies comparing regression systems 10
- Over- and underprediction 13
- Summary of the results of regression studies 20

Correlational comparisons across studies 21
- Median correlations for predominantly white or Anglo samples 21
- Median correlations for black samples 24
- Studies of Mexican American or Chicano samples 26
- Median correlations for minority samples 26
- Minority/nonminority comparisons of median correlations 26
- Summary of correlational comparisons across studies 27

Correlational comparisons within studies 29
- Within-studies comparisons of the high school record as a predictor 29
- Within-studies comparisons of verbal test scores as predictors 29
- Within-studies comparisons of quantitative test scores as predictors 31
- Within-studies comparisons of composite test scores as predictors 31
- Within-studies comparisons of the combination of HSR and tests as predictors 31
- Summary of correlational comparisons within studies 38

Special issues in prediction 39
Are predictive validities declining? . . . 39
Long-range criteria 39
Individual versus group considerations 40
Criterion bias. 41

Internal analyses 42
Comparison of proportions responding correctly to items 42
The item characteristic curve method . . 43
The analysis of variance method. 44
The chi-square method. 44
The distractor response analysis method. 44
Factor analysis methods 45
Analysis of content 45
Summary of internal analysis methods . . 46

Summary and conclusions 47

References. 50

Annotated bibliography 55

Appendix tables

A-1. Correlations between high school grades, high school ranks, and test scores (as predictors) and college GPA (as criterion) for general (unspecified) populations 65

A-2. Correlations between high school grades, high school ranks, and test scores (as predictors) and college performance (as criterion) for white male and female samples 74

A-3. Correlations between high school grades, high school ranks, and test scores (as predictors) and college performance (as criterion) for black (or predominantly black) male and female samples 79

A-4. Correlations between high school grades, high school ranks, and test scores (as predictors) and college performance (as criterion) for Mexican American or Chicano male and female samples 84

A-5. List of abbreviations 86

Tables and figures

Table 1. Predictive studies summary table 8

Table 2. Studies comparing regression systems. 11

Table 3a. Over- and underprediction when minority means entered in common or nonminority equation based on the high school record only 14

Table 3b. Over- and underprediction when minority means entered in common or nonminority equation based on verbal test scores only 15

Table 3c. Over- and underprediction when minority means entered in common or nonminority equation based on quantitative test scores only. 16

Table 3d. Over- and underprediction when minority means entered in nonminority equation based on combined verbal and quantitative test scores 17

Table 4. Over- and underprediction when minority means entered in common or nonminority equation based on combined information from the high school record and test scores. 18

Table 5. Comparison of high school record and test scores as predictors of freshman or cumulative grade-point average for white (or predominantly white) male and female samples 23

Table 6. Comparison of high school record and test scores as predictors of freshman or cumulative grade-point average for black (or predominantly black) male and female samples 24

Table 7. Comparison of high school record and test scores as predictors of freshman or cumulative grade-point average for minority (or predominantly minority) male and female samples 27

Table 8. Studies comparing the high school record as a predictor of college performance for Anglo, black, and Chicano samples. 30

Table 9. Studies comparing verbal test scores as predictors of college performance for Anglo, black, and Chicano samples. 32

Table 10. Studies comparing quantitative test scores as predictors of college performance for Anglo, black, and Chicano samples. 34

Table 11. Studies comparing composite test scores as predictors of college performance for Anglo, black, and Chicano samples. 36

Table 12. Studies comparing multiple correlations of high school record combined with test scores as predictors of college performance for Anglo, black, and Chicano samples. 37

Table 13. Prediction studies using nontraditional methods 40

Figure 1. Percentage eliminated in primary pilot training. 2

Figure 2. An illustration of the effect of selection on the correlation between test scores and a criterion. 3

Figure 3. Prediction of freshman GPA from NMSQT. 4

Figure 4. An illustration of over- and underprediction 5

Preface

Any review of research necessarily requires summarizing many diverse types and pieces of information. In reporting such summaries, it is often necessary to make difficult decisions about how to label data that have been labelled by others in varying ways. Moreover, populations are sometimes classified very broadly, sometimes very specifically, and there is frequently an overlap between one population and another. It may be helpful to describe some perplexities in labelling and classification that were met in the preparation of this report.

The principal populations included in this review are male and female subpopulations of:

- Black, Afro-American, or Negro populations
- Chicano, Mexican American, Spanish-surnamed, or Hispanic populations
- White, Caucasian, or Anglo populations
- Unspecified populations.

The terms black, Chicano, white, and Anglo are used primarily here because they are common ones encountered in the research literature. In each case, the term chosen depends on whether or not groups are being compared and, if so, which ones. Blacks are usually compared to whites and Chicanos to Anglos.

The term, Chicano, is commonly used in California and in research studies, although Mexican Americans from other states often prefer to describe themselves as Mexican Americans. Blacks may also have Spanish surnames and identify with Spanish-speaking populations. The term Anglo is used to describe white populations that are not also Chicano or Mexican American because that is the way the term is used in the research studies reviewed. This usage of the word Anglo is different from the usage that means "of English origin." There are undoubtedly other confusions in labelling, but it is hoped that readers will not be too distracted by them.

Acknowledgments

The extensive summary tables in the appendix were compiled by Shula Minsky, Senior Research Assistant, Educational Testing Service. This required careful reading of numerous research reports and published papers, extracting often elusive pieces of information, and organizing them into an understandable and presentable format. Because the tables were the basis on which several of the analyses presented in the body of the report were made, this was a contribution of considerable significance.

The author also thanks William H. Angoff, Stephen H. Ivens, and Donald A. Rock for their helpful comments on an earlier draft.

Special thanks are due to Chris Rushinski, who typed the manuscript and prepared the designs from which the illustrations were rendered for reproduction.

Abstract

This report reviews and summarizes the data of published reports and papers pertaining to the validity of measures commonly used at college entrance. These measures consist of various types of quantitative indices used to describe the high school record and various types of academic tests. The data were categorized by sex and ethnic classifications where possible, and comparisons were made.

Most of the observations that had been made in previous reviews of this kind were upheld. Both the high school record and standardized academic tests appear to be generally useful at college entrance. Both tend to predict college performance, but they are most appropriately used in combination. For Anglo populations, the high school record usually yields slightly higher predictive correlations. For black populations, however, the data suggest no consistent superiority of the high school record as a predictor of college performance. The amount of data available for Mexican American or Chicano samples was not sufficient for any generalizations about predictive correlations, but the magnitude of those correlations available were on the average lower than those for Anglo populations. For recent years, correlations for black samples tended to be slightly lower than those for Anglo samples.

A primary objective in the review was to consider whether inferences made about different populations from the high school record and test scores tended to be correct inferences. Where inferences tended to be systematically incorrect, there was an interest in the direction of the error. Since the high school record and test scores are used in making inferences about probable college performance, the systematic inferential errors of interest consisted of the degree to which college performance of populations were over- or underpredicted when the data used in making the inference were not limited to data for that population.

Data from a number of studies show that the college performance of black populations, male and female, have been consistently overpredicted by the traditional academic predictors (the high school record and standardized test scores) when predictions are based on data from white or predominantly white samples. The high school record would appear, from the studies reviewed, to be the principal source of this overprediction. Standardized test scores also have tended to overpredict college performance of black populations, but a combination of the high school record and test scores appears to minimize overprediction. In contrast, published data show women to be consistently underpredicted by the traditional predictors when predictions are based on data from male or predominantly male samples.

This review indicates a number of topics where further research appears to be needed. These topics include the study of: long-range criteria,

objective criteria, prediction within major field, the high school record for different populations, college grades for different populations, neglected populations, factorial consistency of measures across populations, item characteristic curves, expert judgment of test content, and the rationale of the content of standardized tests.

Introduction

The high school record and scores on college entrance tests constitute the primary sources of information used in college admissions and placement. The degree to which these measures lead to appropriate inferences about the academic competencies of persons entering college is therefore of considerable importance. The study of measures of these types and the appropriateness of inferences made from them is normally encompassed in the notion of validity. Since validity can be a rather complex topic, it will be useful to review briefly some of its foundations and to indicate current views of it.

Accepted notions of validity

The currently accepted notion of validity was set forth originally by Cronbach and Meehl (1955), and these ideas reflect much of what is contained in the latest version of the American Psychological Association (APA) standards for educational and psychological tests (APA, 1974). In this view, validity is essentially construct validity and evidence for it is accumulated by various means. Anastasi (1976) writes: ". . . construct validation requires the gradual accumulation of information from a variety of sources. Any data throwing light on the nature of the trait under consideration and the conditions affecting its development and manifestations are grist for this validity mill." (p. 151)

Willingham (1976) notes that "often the term *validity* is conceived narrowly—simply as a relationship between a test score and some measure of success in subsequent activity." And, as conceived by Loevinger (1957) as an overarching conceptual framework, construct validity includes content validity, concurrent validity, and predictive validity. Although the approach has had its critics (for example, Bechtoldt, 1959), the adoption of construct validity as described by Cronbach and Meehl, by the APA standards committee, and its general acceptance by recent authors lends credence to its acceptability.

To better understand validity conceived more broadly, it is useful to excerpt from the APA standards statements describing content, criterion-related, and construct validity:

"Evidence of *content validity* is required when the test user wishes to estimate how an individual performs in the universe of situations the test is intended to represent. Content validity is most commonly evaluated for tests of skill or knowledge; it may also be appropriate to inquire into the content validity of personality inventories, behavior checklists, or measures of various aptitudes." (p. 28)

"Criterion-related validities apply when one wishes to infer from a test score an individual's most probable standing on some other variable called a criterion. Statements of predictive validity indicate the extent to which an individual's future level on the criterion can be predicted

from a knowledge of prior test performance." (p. 26)

"Evidence of *construct validity* is not found in a single study; rather judgments of construct validity are based upon an accumulation of research results. In obtaining the information needed to establish construct validity, the investigator begins by formulating hypotheses about the characteristics of those who have high scores on the test in contrast to those who have low scores. Taken together, such hypotheses form at least a tentative theory about the nature of the construct the test is believed to be measuring." (p. 30)

Beyond these fundamental ideas about validity, one must always ask the question: Validity for what? In the case of college entrance, tests are used primarily as a standard for comparing students from a variety of secondary schools where grading standards vary considerably. When used appropriately, the test is used only in combination with the record of high school performance. The test score serves as a balancing factor since it is acknowledged that one school's A may be another school's C. Accordingly, the test contributes to fairness in the evaluation of entrants or potential entrants by recognizing that all secondary schools are not of the same quality nor do their students have the same competencies. Because of these differences in secondary schools, it is recognized that the high school record may not lead to valid inferences when taken alone. Thus, when one considers the validity of college entrance measures, all measures used at college entrance should be evaluated with respect to inferences made from them, either individually or in combination.

Population validity

The concept of population validity has been used in reference to the generalizability of research findings across different populations (Bracht and Glass, 1968; Messick and Barrows, 1972). Messick (1975) suggested that generalizability of the meaning of a measure across groups is an important aspect of construct validity and emphasized that one validates not a specific measure but interpretations of data derived from the measure. It is important that the measure have similar properties and patterns of relationships within different populations. Willingham (1976) used the term population validity in considering those inferences made about the meaning and validity of test scores and other psychometric measures for particular populations. That is, an inference is valid for a population to the extent that it leads to correct judgments about members of that population. An inference may be valid for one population and not for another. Population validation implies an examination of possibly incorrect inferences for each population of interest for each measure or combination of measures used.

Since some of what is included under the rubric of population validity is also discussed under the popular heading, test bias, it is useful to consider how the two terms are related. Test bias discussions usually focus on tests viewed rather narrowly. Most often, test bias deals with differential predictive qualities of tests such as aptitude tests. Test bias does not usually recognize the broader definition of tests used in the APA Standards (APA, 1974), the Equal Employment Opportunity Coordinating Council (EEOCC) Guidelines (EEOCC, 1977), and in other seminal documents where a test is viewed as any quantified measure (including interview ratings, grades, and the like). Some test bias studies consider the internal nature of measures through item analysis studies, but since inferences are not made from responses to items taken singly, there are limitations to this kind of analysis. Discussions of test bias often lead to considerable confusion because there is no agreed upon definition of it (Flaugher, 1978). In contrast, the concept of validity has been well developed and though opinions differ even with respect to validity, the recognized professional organizations have adopted well-articulated

positions. By using the term population validity, one hopes to go beyond the limitations of test bias approaches and to examine measures from a much broader perspective. Essentially, population validity is construct validity for each population of interest considered separately.

There are, of course, numerous subgroups or subpopulations for whom validity might be demonstrated. But the compelling interest is with respect to those groups that may represent a logically possible discontinuity for validity. Women, minorities, older adults, and handicapped people are four such groups, and these it would seem should be the focus at present. As time progresses, other groups may surface as focal points when considering population validity.

The review of population validity contained in this report begins with a discussion of some of the problems encountered in empirical validation, since that is the most common type of validation.

Problems in empirical validation

Before presenting data on empirical validation studies, it is essential to outline a few possible pitfalls in their interpretation. Unfortunately, most often data and analyses are presented without these necessary caveats. Since many of the interpretational problems are related to the nature of correlational analysis, a basic understanding of that is necessary for proper understanding of the following discussion. No general introductory background will be provided here, but useful descriptions of correlational analysis as it is applied in educational and psychological research may be found in a number of texts (e.g., Anastasi, 1976; Cronbach, 1970). Several important topics related to proper understanding of empirical validation studies are discussed in the following sections.

Criterion problems

Since much of empirical validation is concerned with criterion-related validity, it is absolutely essential to have a criterion before it can be conducted. Finding and developing criteria represent one of the most formidable obstacles in educational and psychological research. Because of the difficulties inherent in locating criteria, the criteria used are always somewhat less than desirable. For example, research on anxiety measurement may be conducted by correlating the scores on a proposed measure of anxiety with some other observation that is believed to be related to anxiety (e.g., amount of palm sweat, pulse rate, respiration rate). Given limitations of the criteria such as the possibility that changes in them may have been caused by something else other than the anxiety measure of interest (pulse rate might be increased because of caffeine or nicotine intake), one obviously does not necessarily expect a high correlation. The researcher is pleased, in this kind of research, if there is any statistically significant correlation at all. Usually, the better the criterion, the higher the level of correlation expected for a valid measure—but not always. If a criterion measure is contaminated by some kind of dependence on the instrument being validated, a very high (but spurious) correlation may be obtained (see Gulliksen, 1976; Fricke, 1975, for other reasons to be suspicious of high correlations).

In the study of college entrance tests, the most common criterion used is the grade-point average (GPA) obtained in college. Usually, the GPA for the freshman year is used because it is convenient (data can be collected relatively soon), the curriculum involved is more comparable for students, and there is less attrition in students for the freshman year than is the case for longer-range criteria. It is assumed that the freshman GPA is a reasonably reliable and valid measure of freshman college performance and, additionally, that it is unbiased. Often the reliability or validity of the GPA for a particular sample is questionable. More-

over, a particular problem with the GPA is that it is based on different courses for different students. Since it is well known that different types of students (e.g., males and females) systematically choose different kinds of courses and majors, the relationship expected between a measure and GPA may differ considerably. Measures of mathematical ability, for example, should not be expected to correlate highly with GPA for students who tend not to enroll in mathematics or science courses.

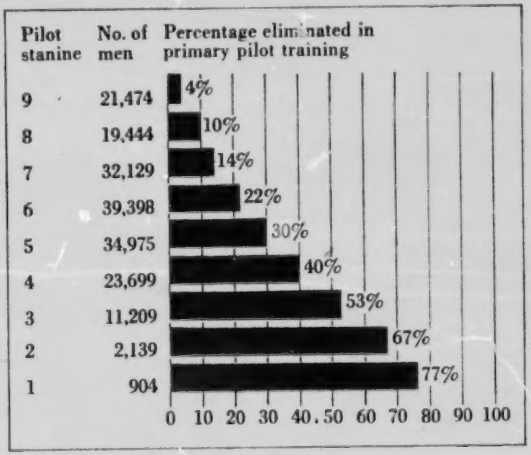
A criterion may also be biased. If systematic error is introduced in the criterion measure by some known or unknown factor, it may be biased. An example might be supervisors' ratings where the supervisors are all males and those being rated are both males and females. Male supervisors may tend to give males higher ratings on traditional male tasks, even though actual performance of the tasks in the situation being rated may actually be as high for females as it is for males. It is difficult to prove or disprove the existence of criterion bias. One way it can be minimized is through double-blind procedures where work samples are evaluated by raters who have no other knowledge or information concerning those being rated and where those being rated have no knowledge or information concerning the raters (see Breland, 1977, for an example of a double-blind essay criterion).

A criterion may be long-range or short-range, and the greater the time interval between the criterion measurement and the measure to which it is being compared the greater the chance for the introduction of extraneous factors that confound validation. Thus, while a popular notion is that college entrance test scores should correlate highly with life success (if they are valid) the probability is that they will not. Such a notion would be similar to a suggestion that tests used to select pilots for pilot training in World War II should have correlated highly with numbers of combat victories, numbers of medals awarded, rank attained, or some other such quantifiable measure of success. Actually, there is little evidence that pilot selection tests did predict success in the real world of war. Despite the lack of this kind of evidence, pilot selection tests are regarded as one of the most successful psychometric experiences in recent history because they correlated well with success in the pilot training program. Because of this relationship between test scores and success in a training program, pilots were produced much more rapidly and at much less expense than would have been possible without effective screening measures. Figure 1 shows the relation between performance on a pilot selection test battery and elimination from primary flight training.

Restriction of range

The pilot selection example is also useful for purposes of illustrating another problem in empirical validation. Despite what appears in Figure 1 to be a rather strong relationship between scores (stanine levels) on the pilot selection battery and success in pilot training, the relationship would

Figure 1



(Adapted from Anastasi, 1976, p. 102.)

have been even greater had all who applied been admitted to pilot training. Note that only 904 men were allowed into the training program who were in the lowest stanine even though many more scored in this range. Apparently, there was other information available that had suggested they might be good risks even though they had low scores. Still, 77 percent of these 904 men were eliminated. Had the total of all applicants been admitted to the training program many more would have been at the lower stanines and, consequently, the relationship observed between test scores and training program success would have been even higher because of the greater variance of the test scores.

The low frequency of pilots at the lower stanine is illustrative of a common phenomenon in validity research known as restriction of range. That is, the range of abilities has been restricted by the selection out of disproportionate numbers of applicants at low stanine levels on the Pilot Selection Battery. Those selected out applicants never entered the training program. As a result, there is no way of knowing how they would have performed in the training program. For validity research the phenomenon of restriction of range is important because it is another source of attenuation in the validity coefficient. A graphical representation of how restriction of range limits correlation coefficients is shown in Figure 2. The shaded areas in the figure represent those selected and thus those for whom criterion data are available. Since the narrowness of an elliptical scatter of points suggests the degree of correlation, the thickness of the shaded portion of the ellipse illustrates how the correlation coefficient would be attenuated by the selection process.

Attenuation resulting from unreliability

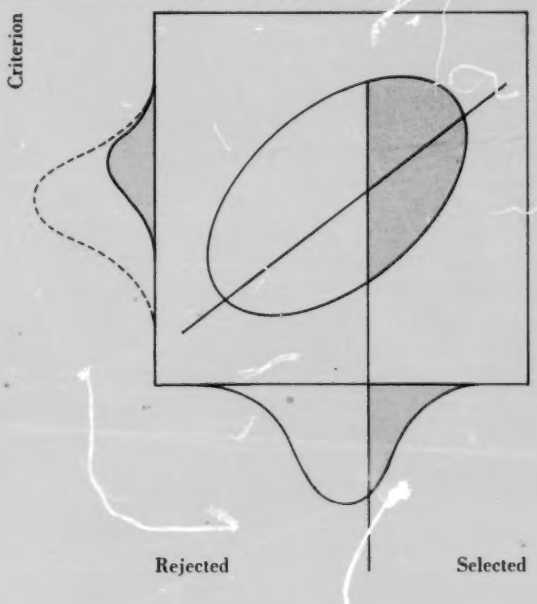
The correlation coefficient is also attenuated by unreliabilities in the measures being correlated. Both the predictor and the criterion measures are imperfect measures, but if one is using the correlation coefficient to evaluate the validity of the predictor it is not appropriate to correct the correlation for unreliability in the predictor (because any unreliability in the predictor should be considered in its evaluation). It is customary, therefore, to correct validity coefficients for errors in the criterion measure only. This may be done by means of the relationship:

$$r'_{xy} = \frac{r_{xy}}{\sqrt{r_{yy}}}, \text{ where } r'_{xy} = \text{the corrected correlation coefficient}$$

r_{xy} = the uncorrected correlation coefficient

r_{yy} = the criterion reliability

Figure 2. An illustration of the effect of selection on the correlation between test scores and a criterion. The validity coefficient within the selected group underestimates the actual predictive value of the test. (Adapted from Stanley and Hopkins, 1972, p. 110.)



Usually, accurate estimates of r_{yy} are not available. As a result, the correction is not made and coefficients are simply interpreted with caution under the assumption that considerable error in the criterion measure exists. Guilford (1965) observes: "Many a validity coefficient reported in the literature is of very uncertain meaning because errors of measurement in the criterion were not taken into account. The reliability of ratings, even of the better ones, is characteristically about .60. With such criterion, validity coefficients are about 25 percent underestimated. Too often the reliability problem of a criterion is entirely ignored." (p. 488)

Suppose the validity coefficient (the correlation between a predictor and a criterion) is .40 and the reliability of the criterion is .60. Then,

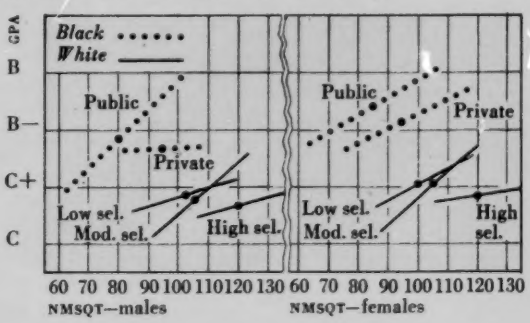
$$r'_{xy} = \frac{.40}{\sqrt{.60}} = \frac{.40}{.77} = .52$$

And this estimate assumes, of course, that the uncorrected correlation, .40, is not also attenuated by restriction in range or other influences.

Problems of pooled samples

A particularly bothersome problem often encountered in empirical validation in academic settings is the attenuation in correlation coefficients caused by pooling data collected at different institutions having widely differing grading standards. The problem is illustrated by the results of a study by Borgen (1972), although numerous other studies have encountered it. Borgen collected data on a large number of black high school seniors nominated by their principals for participation in a national scholarship program. Data included the high school rank (HSR), scores on the National Merit Scholarship Qualifying Test (NMSQT), and college grade-point average. The GPAs were obtained at widely-differing colleges both public and private, predominantly black and predominantly white, highly selective and nonselective, and combinations of these. Figure 3 shows some of Borgen's results as represented by regression lines. Note that, with the exception of males in black private colleges, the slopes of the regression lines are representative of substantial relationships between NMSQT scores and GPA for both males and females. Nevertheless, when data for all students in all types of colleges were pooled, the overall correlation coefficient between NMSQT and GPA was reduced to .03 for males and .06 for females. The attenuation was caused by an inversion in the relationship between NMSQT scores and GPA at black colleges as compared to white colleges. In general, students in the predominantly black colleges had lower NMSQT scores but received higher grades, while the opposite occurred in predominantly white colleges. This kind of apparent paradox has at times led to questioning of the predictive efficacy of test scores for black students in particular (see Clark and Plotkin, 1963; Boyd, 1977). Given this kind of problem, it is best to avoid the use of national samples in validation studies.

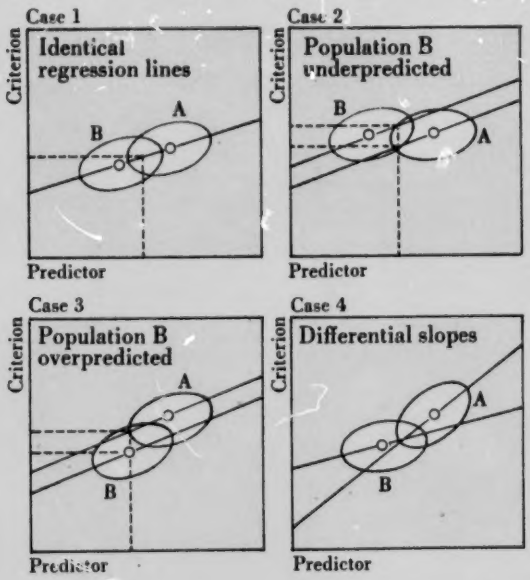
Figure 3. Prediction of freshman GPA *from* NMSQT. *(Solid circles indicate mean* NMSQT *and* GPA; *lines extend 1* SD *above and below the* NMSQT *mean.) (Adapted from Borgen, 1972.)*



Over- and underprediction

Other kinds of spurious inferences can be made when comparing two populations if the regression lines within subsamples are not examined along with the correlation coefficients. The situation is illustrated in Figure 4.

Figure 4. An illustration of over- and underprediction. The ellipses represent the regions within which members of two populations, A and B, fall when predictor measures are plotted against criterion performance. (Adapted from Anastasi, 1976, p. 193.)



In Figure 4, the two ellipses represent the scatter of data from which regression lines are developed for two populations, A and B. In Case 1, the regression lines coincide and, accordingly, any given predictor score would result in the same predicted criterion score regardless of population membership. That is, the same prediction is made for both populations.

In Case 2, however, the regression lines are different even though the correlation coefficients for the two groups may be the same. The regression line for Population B is above and parallel to that for Population A. The Case 2 situation would result in underpredictions for Population B if a regression line were developed from combined data for both populations or if the regression line for Population A were used to make inferences about the probable future performance of members of Population B. In other words, a spurious inference could be made that Population B persons who score low on the predictor will generally perform less well on the criterion measure when in fact they may perform as well as Population A persons, on the average.

Case 3 illustrates a situation essentially opposite to that of Case 2 where the regression line for Population A is above and parallel to that for Population B. Cases 2 and 3 are similar in that Population B has an average predictor score less than that of Population A. The Case 3 situation would result in overpredictions for Population B if a single regression line were developed for combined populations or if a single regression line were used based on the data from Population A. Here the possible false inference is that Population B persons with the same predictor scores as Population A persons will perform as well on the criterion when in fact they do not.

Case 4 shows how over- and underprediction may vary as a function of the predictor score. In this case the slopes of the regression lines are different for the two populations. The result is that Population B would be underpredicted by a Population A regression line for low predictor scores and overpredicted by a Population A regression line for high predictor scores.

Other complexities can also arise in comparing regression lines. The shapes of the ellipses may differ because of a differing scatter of points

about the regression line. And different results (both in regression lines and in correlations) may occur for different predictors (e.g., test scores and grades) taken singly or in combination.

Problems in internal analyses

The interpretational pitfalls in internal analyses (that is, when no external criterion is used) are even greater than are those described above. The methods used—analysis of variance, factor analysis, latent trait theory, for example—are often mathematically and conceptually complex. Some of these problems are outlined in the subsequent section on internal analyses, but a comprehensive examination of this area is beyond the scope of this report.

Summary of problems in empirical validation

In this section a brief review has been made of some common problems in empirical validation before presenting data from validity studies in subsequent sections. Problems of criterion development, criterion bias, problems related to the interval of time between measure and criterion data collection, problems of range restriction, problems caused by unreliability in both measure and criterion, and problems of pooled samples have been noted. Problems specific to the use of the college GPA as a criterion were indicated. One conclusion that follows from these kinds of problems is that extreme caution should be used in the interpretation of correlation coefficients reported in empirical validation studies. Often the magnitude of these correlations is more a function of the quality of the criterion used and the sample studied than of anything else.

Regression comparisons

The most common method for evaluating educational measures has been the study of prediction. At issue are questions of the following type: What is the relationship between the high school record and later performance in college? What is the relationship between a score obtained on a test taken prior to college and college performance? Which of the two—the high school record or the test score—is the better predictor? Is there an advantage in using both predictors in combination? To answer these kinds of questions numerous studies have been conducted by individual colleges, by educational researchers, and by test developers.

With respect to inferences that are made from high school grades and test scores, there are two principal kinds of errors that have been considered in prediction studies. One kind of inferential error is commonly studied through an examination of regression equations. The objective in these regression analyses is usually to determine whether an equation developed for one population, or combinations of populations, is appropriate for other populations. More specifically, the question is whether a given equation over- or underpredicts later academic performance for some particular population. This first kind of prediction study is reviewed in this section. The second kind of prediction study, the correlational study, is reviewed and discussed in the two subsequent sections.

Summary of regression studies

Given the number and variety of regression studies that have been conducted, it is useful to attempt a summarization prior to focusing on specific issues. Table 1 lists 35 studies or reviews of regression studies that have dealt with population validity issues in one way or another. Indicated are the kinds of samples studied, the predictors used, and the criteria used. The code numbers, to the left, will be used in later tabulations to indicate these studies. An annotated bibliography of these studies is included at the end of this report. Explanations of test abbreviations are given in Appendix Table A-5.

With respect to the samples, there are a number of different types of studies represented. Most of the studies have made a comparison of groups, such as blacks vs. whites or Chicanos vs. Anglos. Others have grouped different ethnic classifications and compared minorities vs. nonminorities (Breland, 1977; Wilson, 1978). Very few studies have compared more than two ethnic groups (Goldman and Hewitt, 1976; Scott, 1976). Some of the studies listed in Table 1 have made no comparisons, but they are important because they represent a sampling of blacks or Chicanos that is important for purposes of population validation (Borgen, 1972; Dalton, 1974; Funches, 1967; Hills, 1964; Hills and Stanley, 1970;

Table 1. Predictive studies summary table[a]

Study	Description of samples	Predictors	Criteria
1. Biaggio (1966)	Over 3,000 blacks in three black colleges in Georgia and over 20,000 whites in 15 white colleges	SAT-V, SAT-M	FGPA
2. Borgen (1972)	1,314 black National Merit participants	NMSQT, HSR	FGPA
3. Bowers (1970)	405 blacks (predominantly) and 4,855 whites at the University of Illinois	SCAT-V, SCAT-M, HSR	FGPA
4. Breland (1977)	Minorities and nonminorities in 14 colleges during 1974-75 and four colleges during 1975-76	SAT-V, SAT-M, TSWE, HSR, essays	Freshman English grades, essays
5. Calkins and Whitworth (1974)	1,241 Spanish-surnamed and 1,996 non-Spanish-surnamed at the University of Texas at El Paso	SAT-V, SAT-M, HSR	FGPA
6. Cleary (1968)	Black and white samples from three integrated colleges, two Eastern and one Southwestern	SAT-V, SAT-M, HSR, HSGPA	FGPA, cumulative GPA
7. Cleary, Humphreys, Kendrick, and Wesman (1975)	(Review)		
8. Dalton (1974)	169 blacks, 17 of Mexican-American or Puerto Rican descent, and 15 whites admitted in a special program at Indiana University	SAT-V, SAT-M, HSR	FGPA
9. Davis and Kerner-Hoeg (1971)	Blacks and whites in six integrated colleges in North Carolina	SAT-V, SAT-M, HSR	FGPA
10. Davis and Temp (1971)	Blacks and whites in 19 integrated colleges	SAT	FGPA
11. Farver, Sedlacek, and Brooks (1975)	Freshman blacks entering the University of Maryland in the fall of 1968 and 1969 and randomly sampled white comparison groups	SAT-V, SAT-M, HSGPA	FGPA, Soph GPA, Jnr GPA, Snr GPA, Soph cum, Jnr cum, Snr cum
12. Funches (1967)	540 blacks at Jackson State College	ACT, HSGPA	FGPA
13. Goldman and Hewitt (1975)	Chicanos and Anglos at four universities of a Western multicampus university in 1974	SAT-V, SAT-M, HSGPA	GPA

a. Further details of these studies are contained in the annotated bibliography at the end of this report. Abbreviations are explained in Appendix Table A-5.

Table 1 continued

Study	Description of samples	Predictors	Criteria
14. Goldman and Hewitt (1976)	272 blacks, 188 Chicanos, 852 Orientals and 4,259 whites at UCLA	SAT-V, SAT-M, HSGPA	FGPA
15. Goldman and Richards (1974)	Two entering classes (1971 and 1973) of Chicanos and Anglos at UC Riverside	SAT-V, SAT-M	FGPA
16. Goldman and Widawski (1976)	Blacks and Chicanos at UCLA, UC Davis, UC San Diego, and UC Irvine	SAT, HSGPA	FGPA
17. Harris and Reitzel (1976)	45 blacks and 3,895 whites at the University of North Carolina	SAT-V, SAT-M, HSR	FGPA
18. Hills (1964)	Blacks in three traditionally black colleges in Georgia	SAT-V, SAT-M	FGPA
19. Hills and Stanley (1970)	Blacks in three traditionally black colleges in Georgia	SAT, SCAT, HSGPA	FGPA
20. Kallingal (1971)	225 blacks and 511 randomly selected whites entering Michigan State University in 1968	CQT-V, CQT-N	Soph cum
21. Linn (1973)	(Review)		
22. McKelpin (1965)	Blacks entering North Carolina College at Durham in 1961, 1962, and 1963	SAT-V, SAT-M, HSGPA	FGPA
23. Munday (1965)	Blacks at five different colleges	ACT, HSGPA	FGPA
24. Pandey (1971)	47 black and 33 white graduating seniors	SCAT	Snr cum
25. Peterson (1968)	225 blacks at a college in Georgia	SAT-V, SAT-M	FGPA
26. Pfeifer and Sedlacek (1971)	116 blacks and 178 whites at the University of Maryland	SAT-V, SAT-M, HSGPA	FGPA
27. Scott (1976)	67 blacks, 66 Chicanos, and 878 Anglos in a Southwestern university	ACT, HSGPA	Soph cum
28. Silverman, Barton, and Lyon (1976)	135 blacks, 272 whites, and 121 Jews at Roosevelt University	ACT, HSR	FGPA, Soph GPA
29. Stanley and Porter (1967)	Blacks and whites in 3 black and 14 white colleges in Georgia	SAT-V, SAT-M, HSGPA	FGPA
30. Temp (1971)	Blacks and whites in 13 integrated institutions	SAT-V, SAT-M	FGPA

Table 1 continued

Study	Description of samples	Predictors	Criteria
31. Thomas and Stanley (1969)	736 blacks in a predominantly black university (and a review of other studies)	SAT-V, SAT-M, SCAT-V, SCAT-M, HSR	FGPA
32. Warren (1976)	Chicanos and Anglos in three California state colleges	ACT, SAT, HSGPA	FGPA, Soph GPA, Jnr GPA, Snr GPA
33. Wasserman (1978)	61 blacks and 774 whites in a New Jersey university	SAT-V, SAT-M	Sociology course grades
34. Wilson (1969)	Blacks and whites in four women's liberal arts colleges	SAT-V, SAT-M, HSR, CB Ach avg.	FGPA
35. Wilson (1978)	Minorities and nonminorities in two institutions	SAT-V, SAT-M, HSR, CB Ach avg.	FGPA, Soph GPA, Snr GPA, Soph cum, Snr cum

McKelpin, 1965; Munday, 1965). The study by Borgen was mentioned earlier in the previous section on problems in empirical validation, since it illustrates pitfalls encountered when national samples of blacks, or any other group, are pooled. Two of the references listed are reviews that summarize the results obtained from a number of these same studies (Linn, 1973; Cleary et al., 1975).

The predictors used in the studies of Table 1 include a number of different measures, but they are all measures that have been used in one way or another in the college entrance context. They may not have been used in admission or selection. Only one of the studies used an essay as a predictor (Breland, 1977) and only two used achievement tests (Wilson, 1969 and 1978). Often, but not always, the predictors were combined in multiple regression analyses. Some studies report both zero-order and multiple correlations using the predictors indicated.

Although the freshman grade-point average (FGPA) was the most commonly used criterion in the studies listed in Table 1, other important criteria were also used. Some studies used longer-range criteria such as sophomore, junior, or senior GPA (Farver et al., 1975; Kallingal, 1971; Warren, 1976; Wilson, 1978). Only one study used an essay as a criterion (Breland, 1977).

Studies comparing regression systems

One of the basic kinds of analyses conducted is the comparison of parameters of regression equations for different groups. The three parameters of principal interest are the errors of estimate, the slopes, and the intercepts. In the procedure advocated by Gulliksen and Wilks (1950), parameters are compared in a specified sequence. If significant differences are found at any stage of the sequence, the regression systems are considered to be different. In Table 2, a number of the studies of Table 1 are summarized with respect to their findings of significant differences in errors of estimate, slopes, and intercepts. Table 2 represents an extension of a similar table developed by Linn (1973). Note that all of the studies do not report tests of significance for all of the parameters. In the Gulliksen-Wilks procedure, the er-

Table 2. Studies comparing regression systems

Study code/ comparison[b]	Groups compared[a]	Significant differences		
		Errors of estimate	Slopes	Intercepts
3	Blacks and whites	–	Yes	–
4	Minorities and nonminorities	–	No	Yes
5	Spanish surnamed and non-Spanish surnamed	–	No	Yes
6/1	Blacks and whites	–	No	No
6/2	Blacks and whites	–	No	No
6/3	Blacks and whites	–	No	Yes
9/1	Blacks and whites	–	No	Yes
9/2	Blacks and whites	–	No	No
9/3	Blacks and whites	–	No	Yes
9/4	Blacks and whites	–	No	No
9/5	Blacks and whites	–	No	No
9/6	Blacks and whites	–	Yes	–
11	Blacks and whites	–	–	–
13/1	Chicanos and Anglos	–	Yes	–
13/2	Chicanos and Anglos	–	No	–
13/3	Chicanos and Anglos	–	Yes	–
13/4	Chicanos and Anglos	–	Yes	–
14/1	Blacks, Chicanos, Anglos, and Orientals	–	No	Yes
14/2	Blacks, Chicanos, Anglos, and Orientals	–	No	No
14/3	Blacks, Chicanos, Anglos, and Orientals	–	No	No
14/4	Blacks, Chicanos, Anglos, and Orientals	–	Yes	No
15/1	Chicanos and Anglos	–	Yes	–
15/2	Chicanos and Anglos	–	Yes	–
17	Blacks and whites	–	–	–
20	Blacks and whites	–	–	–
26	Blacks and whites	–	–	–
27/1	Chicanos and Anglos	No	No	No
27/2	Blacks and whites	No	No	No
27/3	Blacks and Chicanos	No	No	No
28	Blacks, whites, and Jews	–	–	–
30/1	Blacks and whites	No	No	Yes
30/2	Blacks and whites	No	No	Yes
30/3	Blacks and whites	Yes	–	–
30/4	Blacks and whites	No	No	Yes
30/5	Blacks and whites	Yes	–	–
30/6	Blacks and whites	No	No	No
30/7	Blacks and whites	No	No	Yes
30/8	Blacks and whites	No	No	No
30/9	Blacks and whites	Yes	–	–

Table 2 continued

Study code/ comparison[b]	Groups compared[a]	Significant differences: Errors of estimate	Slopes	Intercepts
30/10	Blacks and whites	Yes	–	–
30/11	Blacks and whites	No	Yes	–
30/12	Blacks and whites	No	No	No
30/13	Blacks and whites	Yes	–	–
32/1	Chicano and Anglo men	No	Yes	–
32/2	Chicano and Anglo men	No	No	No
32/3	Chicano and Anglo women	No	No	No
32/4	Chicano and Anglo women	No	No	No
32/5	Chicanos and Anglos	No	Yes	–
32/6	Chicanos and Anglos	No	No	No
32/7	Chicano and Anglo men	Yes	–	–
32/8	Chicano and Anglo women	No	No	No
32/9	Chicanos and Anglos	No	No	No
32/10	Chicano and Anglo men (Soph GPA)	No	No	No
32/11	Chicano and Anglo women (Soph GPA)	No	No	No
32/12	Chicanos and Anglos (Soph GPA)	No	No	No
32/13	Chicanos and Anglos (Jnr GPA)	No	No	No
32/14	Chicanos and Anglos (Jnr GPA)	No	No	No
32/15	Chicanos and Anglos (Snr GPA)	No	Yes	–
32/16	Chicanos and Anglos (Snr GPA)	No	No	Yes
32/17	Chicanos and Anglos (Upper div GPA)	No	Yes	–
35/1	Blacks and whites	No	No	Yes
35/2	Blacks and whites (Jnr cum GPA)	Yes	No	Yes
35/3	Blacks and whites	No	No	No
35/4	Blacks and whites (Snr cum GPA)	Yes	No	No
Summary of significant differences		8 yes 29 no	13 yes 40 no	13 yes 27 no
Summary of significant differences of at least one type where blacks are involved				19 yes 6 no
Summary of significant differences of at least one type where Chicanos are involved				10 yes 14 no
Summary of significant differences of at least one type				33 yes 25 no

a. Unless otherwise noted, the criterion used was freshman grades.

b. The study code number, given first, refers to Table 1 where studies are summarized. The second code simply indicates a different comparison within the same study.

rors of estimate for the groups are first compared. If significant differences are found, differences in slopes and intercepts are not tested. If errors of estimate are not different, then differences in slopes are considered. If significant differences in slopes are found, differences in intercepts are not considered. Not all of the studies of Table 2 used this procedure, however.

At the bottom of Table 2 a summary is made of the cumulative findings of these studies comparing regression systems. A total of 33 comparisons showed significant differences of at least one type and 25 comparisons showed no significant differences in any of the three parameters. This kind of summary is misleading in one sense. While it suggests that approximately equivalent numbers of comparisons have found regression systems the same as have found them different, it should be recognized that an hypothesis of equivalent regression systems is unlikely to be rejected if the sample sizes studied were small. In some cases of Table 2, the samples studied were relatively small and, therefore, it would be anticipated that some of the 25 comparisons finding no significant differences would have reached the opposite conclusion if larger samples had been compared. Thus, an informed judgment would lead one to suspect that, generally, regression systems for different groups are not the same.

For the comparisons involving blacks primarily, a total of 19 resulted in significant differences of at least one kind, while only 6 comparisons indicated no differences between blacks and either Anglos or Chicanos. Studies comparing Chicanos primarily with other groups resulted in 10 comparisons with significant differences of one type and 14 comparisons suggesting no differences. Again, the difficulty of rejecting the null hypothesis must be noted, but the cumulative results available comparing Chicano and Anglo regression systems are not demonstrative of differences. Goldman and Hewitt (1976) emphasized, moreover, that even when they found significant differences between Chicanos and Anglos in regression systems, the variance accounted for by the differences was so small as to make differences observed of no practical importance.

Over- and underprediction

Perhaps of greater interest than comparisons of statistical significance are comparisons of the outcomes that would have occurred if a regression system derived from the data of one group were applied to the predictors for another group. Typically, the minority predictors have been used in a nonminority equation to determine whether the nonminority equation would be inappropriate if used for minorities. In Tables 3a, 3b, 3c, 3d, and 4, some of the studies of Table 1 are summarized with respect to predicted and actual performance of minority groups. The difference indicated in the last column of these tables is the difference between predicted and actual performance based on data presented in the study. A positive difference represents overprediction and a negative difference, underprediction. The meaning of overprediction is that the nonminority regression equation predicted that minority students would have performed better in college than they did in fact perform. Underprediction means the opposite.

Table 3a shows a comparison of predicted and actual performance when the predicted performance is based on the high school record only. For those studies reporting a zero-order correlation between the high school record and college performance, the standard deviations for both predictor and criterion, and the average predicted performance, it was possible to compute a prediction equation based on the high school record only. If the average actual criterion performance was also reported, then a comparison could be made between predicted performance (using the high school record only) and actual performance. Table 3a shows that 25 of the 26 comparisons possible indicated an *(Text continued on page 19.)*

Table 3a. Over- and underprediction when minority means entered in common or nonminority equation based on the high school record only

Study code/ comparison	Group predicted[a]	Performance		Difference[b]
		Predicted	Actual	
3/1	Black men	3.14	2.88	+.26 (.17)
3/2	Black women	3.47	3.01	+.46 (.31)
5/1	Spanish-surnamed men	1.68	1.68	.00
5/2	Spanish-surnamed women	2.26	2.09	+.17
6/1	Blacks	1.98	1.80	+.18
6/2	Blacks (using high school rank as predictor)	2.79	1.81	+.98
6/3	Blacks (using GPA as predictor)	2.56	1.81	+.75
9/1	Blacks	2.45	1.76	+.69
9/2	Blacks	2.64	1.92	+.72
9/3	Blacks	2.11	1.52	+.58
9/4	Blacks	2.15	1.89	+.26
9/5	Blacks	2.51	1.77	+.74
9/6	Blacks	2.26	1.66	+.60
13/1	Chicanos	2.74	2.62	+.12
13/2	Chicanos	3.04	2.57	+.47
13/3	Chicanos	2.82	2.62	+.19
13/4	Chicanos	2.89	2.75	+.14
14/1	Blacks	2.72	2.52	+.20
14/2	Chicanos	2.77	2.65	+.12
17/1	Blacks	2.26	1.84	+.42
26/1	Black men	1.83	1.64	+.19
26/2	Black women	2.12	1.75	+.37
35/1	Minorities	9.84	8.74	+1.10 (.29)
35/2	Minorities	2.61	2.54	+.07
35/3	Minorities (using Sem 8 GPA as criterion)	10.98	9.86	+1.28 (.34)
35/4	Minorities (using Sem 8 GPA as criterion)	3.27	3.01	+.26

Summary:	
Median difference	+.28
Total overpredictions	25 +'s
Total underpredictions	0 −'s

a. Unless otherwise noted, the criterion used was freshman grades.

b. Differences in parentheses are rough estimates of equivalent differences for a 0 to 4 grading scale for those comparisons where a different grading scale was used.

Table 3b. Over- and underprediction when minority means entered in common or nonminority equation based on verbal test scores only

Study code/ comparison	Group predicted[a]	Performance		Difference[b]
		Predicted	Actual	
3/1	Black men	3.29	3.15	+.14 (.09)
3/2	Black women	3.45	3.09	+.36 (.24)
4/1	Minorities	6.02	5.36	+.66 (.19)
5/1	Spanish-surnamed men	1.52	1.68	−.16
5/2	Spanish-surnamed women	1.92	2.09	−.17
6/1	Blacks	2.02	1.82	+.20
6/2	Blacks	1.88	1.80	+.08
6/3	Blacks	2.06	1.81	+.25
9/1	Blacks	2.30	1.76	+.54
9/2	Blacks	2.04	1.92	+.12
9/3	Blacks	1.97	1.52	+.45
9/4	Blacks	1.79	1.89	−.10
9/5	Blacks	1.89	1.77	+.12
9/6	Blacks	1.93	1.66	+.27
13/1	Chicanos	2.72	2.62	+.10
13/2	Chicanos	2.74	2.57	+.17
13/3	Chicanos	2.84	2.62	+.22
13/4	Chicanos	2.82	2.75	+.07
14/1	Blacks	2.68	2.52	+.16
14/2	Chicanos	2.73	2.65	+.08
15/1	Chicanos	2.52	2.28	+.24
15/2	Chicanos	2.70	2.58	+.12
20/1	Blacks	2.44	2.27	+.33
26/1	Black men	1.71	1.64	+.07
26/2	Black women	2.07	1.75	+.32
35/1	Minorities	9.30	8.74	+.56 (.15)
35/2	Minorities	2.79	2.54	+.25
35/3	Minorities (using Sem 8 GPA as criterion)	10.29	9.89	+.40 (.11)
35/4	Minorities (using Sem 8 GPA as criterion)	2.89	2.72	+.17

Summary:	
Median difference	+.16
Total overpredictions	26 +'s
Total underpredictions	3 −'s

a. Unless otherwise noted, the criterion used was freshman grades.

b. Differences in parentheses are rough estimates of equivalent differences for a 0 to 4 grading scale for those comparisons where a different grading scale was used.

Table 3c. Over- and underprediction when minority means entered in common or nonminority equation based on quantitative test scores only

Study code/ comparison	Group predicted[a]	Performance		Difference[b]
		Predicted	Actual	
3/1	Black men	3.02	3.15	−.13 (.09)
3/1	Black women	3.17	3.09	+.08 (.05)
5/1	Spanish-surnamed men	1.54	1.68	−.14
5/2	Spanish-surnamed women	2.02	2.09	−.07
6/1	Blacks	2.05	1.82	+.23
6/2	Blacks	1.80	1.80	.00
6/3	Blacks	2.14	1.81	+.33
9/1	Blacks	2.26	1.76	+.50
9/2	Blacks	2.15	1.92	+.23
9/3	Blacks	2.02	1.52	+.50
9/4	Blacks	1.75	1.89	+.14
9/5	Blacks	1.94	1.77	+.17
9/6	Blacks	1.97	1.66	+.31
13/1	Chicanos	2.76	2.62	+.14
13/2	Chicanos	2.77	2.57	+.20
13/3	Chicanos	2.88	2.62	+.26
13/4	Chicanos	2.89	2.75	+.14
14/1	Blacks	2.72	2.52	+.20
14/2	Chicanos	2.78	2.65	+.13
15/1	Chicanos	2.55	2.28	+.27
15/2	Chicanos	2.72	2.58	+.14
20/1	Blacks	2.56	2.27	+.29
26/1	Black men	1.65	1.64	+.01
26/2	Black women	2.07	1.75	+.32
35/1	Minorities	9.39	8.77	+.62 (.16)
35/2	Minorities	2.59	2.54	+.05
35/3	Minorities (using sem 8 GPA as criterion)	10.67	9.86	+.81 (.21)
35/4	Minorities (using sem 8 GPA as criterion)	2.52	2.45	+.05

Summary:	
Median difference	+.16
Total overpredictions	24 +'s
Total underpredictions	3 −'s

a. Unless otherwise noted, the criterion used was freshman grades.

b. Differences in parentheses are rough estimates of equivalent differences for a 0 to 4 grading scale for those comparisons where a different grading scale was used.

Table 3d. Over- and underprediction when minority means entered in nonminority equation based on combined verbal and quantitative test scores

Study code/ comparison	Group predicted[a]	Mean performance		Difference[b]
		Predicted	Actual	
5/1	Spanish-surnamed men	1.47	1.68	−.21
5/2	Spanish-surnamed women	1.90	2.09	−.19
6/1	Blacks	1.98	1.82	+.16
6/2	Blacks	1.81	1.80	+.01
6/3	Blacks	1.94	1.81	+.13
15/1	Chicanos	2.66	2.28	+.38
15/2	Chicanos	2.64	2.58	+.06
17/1	Blacks	1.87	1.84	+.03
30/1	Blacks	2.77	2.58	+.19
30/2	Blacks	2.74	2.30	+.44
30/3	Blacks	2.53	2.26	+.27
30/4	Blacks	2.35	2.14	+.21
30/5	Blacks	2.43	2.06	+.37
30/6	Blacks	2.33	2.17	+.16
30/7	Blacks	2.60	2.15	+.45
30/8	Blacks	2.47	2.45	+.02
30/9	Blacks	2.31	2.11	+.20
30/10	Blacks	2.43	1.96	+.47
30/11	Blacks	2.35	2.30	+.05
30/12	Blacks	2.10	2.11	−.01
30/13	Blacks	4.29	4.05	+.24 (.12)
Summary:				
Median difference				+.16
Total overpredictions				18 +'s
Total underpredictions				3 −'s

a. The criterion used was freshman grades.

b. Differences in parentheses are rough estimates of equivalent differences for a 0 to 4 grading scale for those comparisons where a different grading scale was used.

Table 4. Over- and underprediction when minority means entered in common or nonminority equation based on combined information from the high school record and test scores

Study code/ comparison	Group predicted[a]	Performance		
		Predicted	Actual	Difference[b]
3/1	Black men	2.84	3.15	−.31 (.21)
3/2	Black women	3.25	3.09	+.17 (.11)
5/1	Spanish-surnamed men	1.64	1.68	−.04
5/2	Spanish-surnamed women	1.98	2.09	−.13
6/1	Blacks (using HSR as predictor)	1.86	1.84	+.02
6/2	Blacks (using HSR as predictor)	2.44	1.96	+.48
6/3	Blacks (using HSGPA as predictor)	2.23	1.82	+.41
9/1	Blacks	1.96	1.75	+.21
9/2	Blacks	2.12	1.92	+.20
9/3	Blacks	1.69	1.51	+.18
9/4	Blacks	1.93	1.87	+.06
9/5	Blacks	1.93	1.78	+.15
9/6	Blacks	2.00	1.65	+.35
13/1	Chicanos	2.61	2.62	−.01
13/2	Chicanos	2.60	2.58	+.02
13/3	Chicanos	2.66	2.63	+.03
13/4	Chicanos	2.70	2.75	−.05
14/1	Blacks	2.55	2.51	+.04
14/2	Chicanos	2.67	2.60	+.07
14/3	Blacks	2.52	2.52	.00
14/4	Chicanos	2.69	2.69	.00
14/5	Blacks	2.70	2.67	+.03
14/6	Chicanos	2.62	2.69	−.07
14/7	Blacks	2.52	2.52	.00
14/8	Chicanos	2.61	2.65	−.04
17/1	Blacks	2.04	1.84	+.20
26/1	Black men	1.68	1.64	+.04
26/2	Black women	2.06	1.75	+.31
28/1	Blacks	1.99	1.63	+.36
28/2	Blacks (using Sophomore GPA criterion)	2.22	1.73	+.49
34/1	Blacks	6.75	6.51	+.24 (.06)
34/2	Blacks	2.49	2.46	+.03
34/3	Blacks	2.17	2.11	+.06
34/4	Blacks	2.33	2.35	−.02
35/1	Minorities	8.95	8.74	+.21 (.06)

Table 4 continued

Study code/ comparison	Group predicted[a]	Performance		Difference[b]
		Predicted	Actual	
35/2	Minorities (using 3-yr. cumulative GPA criterion)	9.29	9.02	+.27 (.07)
Summary:				
Median difference				+.05
Total overpredictions				25 +'s
Total underpredictions				8 −'s

a. Unless otherwise noted, the criterion used was freshman grades.

b. Differences in parentheses are rough estimates of equivalent differences for a 0 to 4 grading scale for those comparisons where a different grading scale was used.

overprediction when high school grades were used alone, and one prediction was "perfect." The median overprediction was .28.

Table 3b gives a comparison of predicted and actual college performance when the predicted performance is based on verbal test scores (such as SAT-V) only. Out of the 29 comparisons, only three yielded an underprediction for the minority group examined. The remaining comparisons showed overpredictions, and the median difference between predicted and actual performance was +.16.

Table 3c gives a comparison of predicted and actual college performance when the predicted performance is based on quantitative test scores (such as SAT-M) only. Out of 28 comparisons, only three resulted in underpredictions and 24 resulted in overpredictions. The median difference between predicted and actual performance was +.16, the same as the difference that was obtained for verbal scores.

Table 3d is based on data presented in studies that allow for a comparison of predicted and actual performance when the predicted performance is based on a combination of verbal and quantitative test scores (such as the sum of SAT-V and SAT-M). In 18 of the 21 comparisons an overprediction for the minority group involved resulted, and the median difference between predicted and actual performance was +.16.

It should be expected that the best predictions would occur when information in the high school record is combined with information from tests. Table 4 shows an analysis of over- and underprediction when this is done. Out of the 33 comparisons made, 25 were overpredictions and eight were underpredictions. However, the median overprediction was only +.05, and this is somewhat less than the median overpredictions based on either the high school record alone or on test scores alone.

An interesting contrast with these minority/nonminority over- and underprediction studies is provided by similar analyses that have been conducted comparing men and women. While minorities tend to be overpredicted, women appear to be consistently underpredicted. Linn (1973) reviewed data from 10 institutions showing that in every case women actually performed better in college than would have been indicated by regression equations developed from male samples. The ACT (1973) presented data from 19 colleges where ACT scores and high school grades were used to predict first-semester college grade-point aver-

ages. When a single regression equation developed from combined samples of men and women was used, the performance of women in the first semester of college was almost always underpredicted. Conversely, the same equations almost always overpredicted the performance of men. The consistency of underpredictions for women can be seen also in the data of several of the Table 1 studies, although these analyses are not presented here. No analysis is available of the relative contributions of different predictors to the underpredictions for women, but it has been suggested by Wild (1977) that these underpredictions would be minimized if combinations of predictors were used for the prediction of performance in specific major field areas.

Summary of the results of regression studies

From the review of studies in this section it is clear that regression systems differ systematically between minorities and nonminorities and between men and women. The performance of minorities in college has been consistently overpredicted by equations based on either nonminority samples or combined samples including both minorities and nonminorities. The performance of women has been consistently underpredicted by regression equations based on male samples or on combined samples of men and women. The overpredictions for minorities occur whether the prediction is made from the high school record alone, verbal test scores alone, quantitative test scores alone, or from combinations of grades and test scores. However, the overpredictions for minorities appear to be substantially reduced, on the average, when the high school record and test scores are used in combination as predictors. No comparable analysis is available for women, but it has been suggested that underpredictions for women may also be minimized by the use of combinations of predictors—especially when prediction is limited to performance within specific major fields.

Correlational comparisons across studies

Although the specific analyses presented in predictive studies vary considerably, most report correlation coefficients of one kind or another. Therefore, despite the interpretive problems often encountered with individual correlation coefficients, they serve as a useful medium with which to summarize results from a large number of studies. In Appendix Tables A-1, A-2, A-3, and A-4, reported studies have been listed for four different major populations: those unspecified with respect to race (but predominantly white), whites, blacks, and Chicanos. Male and female correlations are reported within each of the appendix tables as well as correlations based on combined sex groups. Each subsample within each study is indicated along with the correlations reported between college performance and test scores, high school record (either rank or grade point average), and multiple correlations based upon combinations of high school record and test scores. It is possible from these appendix tabulations to make comparisons of the relationships generally obtained for the different predictors for each of the populations of interest. These comparisons can be made by computing median correlations for all reported studies or by considering only comparisons reported within studies. Since many studies do not make group comparisons (because only one kind of group was being studied), the most comprehensive summary is obtained by computing the median across all studies for each group. But because the conditions vary across studies, it is also of interest to consider correlational comparisons made within studies. The across-studies correlations are summarized in this section and the correlations within studies in the subsequent section.

Median correlations for predominantly white or Anglo samples

To provide a basis for comparison with other groups, Table 5 summarizes the correlations that have been obtained in 41 of the 44 studies of Appendix A-1 (samples unspecified with respect to race, but predominantly white) and 17 of the 18 studies of white samples tabulated in Appendix A-2. Studies reporting median correlations are excluded, but are reserved for comparison with the medians computed across studies. These studies reporting median correlations are Ford and Campos (1977), Schrader (1971), Zimmerman (1965), and Biaggio (1966).

In considering Table 5, it is important to emphasize that precise comparisons among groups or among specific predictors is not possible because not only different kinds of groups are involved but different numbers of groups. The summary is useful, however, as a means of generalizing, in an approximate way, about the studies that have been reported. Observe, at the top of Table

5, that there are usually fewer studies of combined sex groups than of males or females considered separately. Exceptions occur, however, for studies reporting predictive correlations for verbal and quantitative test scores. The most sample correlations (206) are multiple correlations of high school record (HSR) and test scores combined as predictors of college performance.

The range of correlations obtained varies enormously and extends from one −.10 correlation obtained between SAT-M scores and college performance for females to a maximum figure of .84 obtained using the high school record as a predictor. The minimum figure was reported in a study by Nichols (1966) of National Merit Scholarship applicants. This extremely low correlation is illustrative of the problem of pooling national samples (see Borgen, 1972) described earlier. The maximum figure was obtained in a study by Endler and Steinberg (1963) of liberal arts students at a newly-opened college in Canada. While statistically significant, this maximum value is based on a sample of only 24 female students. One of the most interesting aspects of the ranges of correlations reported in these studies of predominantly white samples is the tendency for the reported ranges to be smaller when the high school record and test scores are used in a multiple correlation to predict college performance. What is suggested is that when either test scores or high school record is used alone, the correlations obtained can vary somewhat more than when these predictors are combined.

The most meaningful way to compare such correlations is to compute the median and these are given at the bottom of Table 5. The figures in parentheses are averages of the median correlations reported by Ford and Campos (1977) in a summary of College Board member institution data. The Ford and Campos figures serve as a useful reference point because they represent much more control of samples than is possible from a summary of reported studies. The median correlations computed across studies have some interesting features, however. First, in all three cases (males, females, and combined samples), the high school record yields a higher median correlation than any of the test scores. Second, the quantitative test scores tend to result in the lowest median correlations across studies. Third, the combination of high school record and test scores provides the highest median correlation in every case. Fourth, females tend to yield higher median correlations than males. These median correlations are generally of the same pattern as those obtained by Ford and Campos. One notable difference is that the median correlations for the combined sex groups in the Ford and Campos study tend to fall between the male and female samples (as one would expect) while the median correlations computed across studies do not. Obviously, the sample of studies in which correlations for combined sex groups were reported is not directly comparable to the sample of studies reporting correlations for males and females separately.

The Table 5 median correlations may also be compared to median correlations reported by Biaggio (1966) and Schrader (1971). The Biaggio medians are based on 15 public colleges within the state of Georgia. For over 25,000 students Biaggio computed median correlations between test scores and college GPA for four different years extending between 1959 and 1963. For males, using the SAT-V, the median correlations obtained were .39, .41, .34, and .38—for an average of .38. For females, using the SAT-V, the median correlations obtained were .55, .54, .51, and .46—an average of .51. Therefore, the suggestion from Table 5—that female college performance is more predictable from verbal test scores than male college performance—is corroborated by the SAT-V data from Biaggio. For the SAT-M, Biaggio obtained median correlations for males of .41, .37, .38, and .37—an average of .38. For females, on the SAT-M, median correlations of .53, .49, .51,

and .45 were obtained—with an average of .50. Thus, the figures of Table 5, suggesting that not only are females better predicted with verbal scores but also with quantitative scores, are supported by Biaggio's data.

Consider now the median correlations reported in Schrader's (1971) summary for comparison with Table 5. Median correlations were reported for verbal and quantitative test scores, for high school rank, and for combinations of these as predictors of college performance. For the SAT-V, Schrader reported that males have a median correlation of .33 and females .41. For the SAT-M, Schrader reports a median correlation of .30 for males and .36 for females. Again the supposition that females yield higher predictive correlations with either verbal or quantitative test scores is consistent with these data. For the high school record as a predictor of college performance, Schrader reports median correlations of .47 for males and .57 for females; again this is in agreement with Table 5. Schrader reports median multiple correlations, combining high school record and test scores of .55 for males and .62 for fe-

Table 5. Comparison of high school record and test scores as predictors of freshman or cumulative grade-point average for white (or predominantly white) male and female samples

Predictor	Males	Females	Combined sex samples	All samples
Number of samples				
High school record	54	53	35	142
Verbal test scores	30	30	38	98
Quantitative test scores	34	34	36	104
Composite test scores	54	51	31	136
HSR and tests	88	83	35	206
Range of correlations				
High school record	−.14–.72	−.07–.84	.03–.71	−.07–.84
Verbal test scores	−.01–.59	.01–.55	.16–.64	−.01–.64
Quantitative test scores	−.04–.47	−.10–.55	.15–.47	−.04–.57
Composite test scores	.10–.61	−.03–.73	.15–.55	−.03–.73
HSR and tests	.13–.75	.23–.73	.31–.79	.13–.79
Median correlations				
High school record	.47 (.46)[a]	.51 (.52)	.45 (.52)	.48 (.50)
Verbal test scores	.35 (.31)	.44 (.40)	.37 (.37)	.38 (.36)
Quantitative test scores	.36 (.30)	.40 (.38)	.30 (.31)	.35 (.34)
Composite test scores	.38	.39	.42	.39
HSR and tests	.55 (.53)	.56 (.60)	.51 (.58)	.54 (.57)

a. The figures in parentheses are the averages of the medians reported by Ford and Campos (1977) and are offered for comparison because they probably represent better sampling than the other medians based only on published reports.

Note: See Appendix Tables A-1 and A-2 for a listing of the studies on which this table is based.

males which are not only consistent with Table 5, but identical in the case of the .55 figure for males.

In summary, a number of conclusions may be made about the nature of the prediction of college performance for predominantly white male and female populations:

- White females tend to yield higher correlations than do white males when the high school record, verbal test scores, or quantitative test scores are used alone.
- For both white males and white females the combination of test scores and grades provides a higher correlation, generally, than either tests or grades used alone as predictors.
- Considering single predictors independently for white populations, the high school record tends to yield the highest correlations with college performance, verbal test scores the second highest, and quantitative test scores the third highest.

Median correlations for black samples

Table 6 presents an analysis of median correlations obtained across studies that is similar to the analysis presented previously in Table 5 for white samples. The number of black samples available is, of course, somewhat smaller than the number of samples available for the analyses of Table 5. Correlations for 42 different samples of blacks

Table 6. Comparison of high school record and test scores as predictors of freshman or cumulative grade point average for black (or predominantly black) male and female samples

Predictor	Males	Females	Combined sex samples	All samples
Number of samples				
High school record	22	20	28	70
Verbal test scores	12	11	24	47
Quantitative test scores	12	11	24	47
Composite test scores	8	8	26	42
HSR and tests	43	43	24	110
Range of correlations				
High school record	.11–.63	.25–.66	.02–.61	.02–.66
Verbal test scores	.23–.59	.17–.51	.06–.65	.06–.59
Quantitative test scores	.14–.47	.11–.55	.01–.55	.01–.55
Composite test scores	–.32–.54	.27–.64	–.04–.65	–.32–.65
HSR and tests	.30–.78	.29–.72	.31–.70	.29–.70
Median correlations				
High school record	.42	.57	.38	.46
Verbal test scores	.42	.43	.27	.41
Quantitative test scores	.40	.41	.20	.34
Composite test scores	.41	.52	.30	.39
HSR and tests	.57	.64	.48	.57

Note: See Appendix Table A-3 for a listing of the studies upon which this table is based.

were available in which composite test scores were reported. However, correlations based on 110 different black or predominantly black samples were reported in which the combination of high school record and test scores was used for the prediction of college performance.

The range of correlations obtained in the studies of black samples is similar to that obtained for white samples. The trend toward smaller ranges of correlations when the high school record is combined with tests is quite similar to that obtained for the white samples of Table 5. This is especially apparent for the lower correlations which appear to be eliminated when the combination of predictors is used. The maximum correlation is increased slightly to around .70 on the average and the minimum is pulled up to an average of about .30 for males, females, and combined sex samples.

The median correlations for black samples in Table 6 are similar to the median correlations for whites given previously in Table 5. As for the white samples of Table 5, the high school record appears to be a better predictor of college performance for females than for males, but no superiority of high school record (over tests) as a predictor is suggested for the black male samples. The median correlation for black females based on the high school record (.57) seems somewhat higher than the median correlations based on either verbal or quantitative test scores (.43 and .41 respectively), but no substantial difference appears to occur for black males. However, no conclusive statements are warranted in view of the limited numbers of correlations used to compute these medians. There appears to be no ambiguity with respect to the advantage of predictions based on a combination of high school record and test scores. Whether black males, black females, or combined sex groups are considered, the median predictive correlation using a combination of high school record and test scores is substantially higher than that obtained when either test scores or high school record alone is used. For the black male samples, the median correlation using test scores is increased by .15 (from .42 to .57) when test scores are added to the high school record in making the prediction. It is of interest to summarize the increases in median correlations obtained as a result of the addition of test scores to the high school record and to compare this summary for black samples with that made previously for white samples:

	Increases in median correlations resulting from addition of test scores to high school record					
	Whites			Blacks		
Study	M	F	C	M	F	C
The present study	.08	.05	.06	.15	.07	.10
Ford and Campos (1977)	.07	.07	.06			
Schrader (1971)	.08	.05				

This summary table suggests that black female samples are similar to white samples with respect to gains in the prediction of college performance obtained by adding test scores to high school record. Black males, on the other hand, appear to have considerably higher predictive correlations when the test score is added to high school record in the prediction.

The following summary statements seem appropriate for the black samples:

- Black females tend to yield higher correlations than black males whether the high school record or combinations of high school record and test scores are used as predictors.
- For both black males and black females the combination of high school record and test scores as predictors yields a higher median correlation than either used alone.
- For black males, unlike white males and females and black females, the correlational evidence does not suggest that the high school record alone is a better predictor of college performance than test scores used alone.

Studies of Mexican American or Chicano samples

A total of only seven studies of Chicano or Mexican American samples was located for this review and these are summarized in Appendix Table A-4. Because of the limited number of studies it is not possible to develop meaningful tables of median correlations for Chicanos that are comparable to those developed for whites and blacks in Tables 5 and 6. For composite test scores, only four samples are available from which to compute a median. One of these four samples is a male sample, one a female sample, and the other two are combined sex samples. Only nine sample correlations have been reported between verbal and quantitative test scores and college performance. Therefore, no attempt is made here to summarize the results of studies of Chicano populations separate from other populations. In the following section, however, the results of studies of Chicano samples are merged with the results of studies of black samples.

Median correlations for minority samples

Because of the exceedingly limited number of studies on Chicano samples it is interesting to combine the Chicano samples with the black samples and in this way obtain a representative summary for these two minority populations taken together. In Table 7 the correlations for black samples summarized in Table 6 are combined with the correlations for Chicano samples. The effect of this combination is to produce only limited changes in ranges and median correlations from those of Table 6.

As in the case of the white samples, the negative correlations disappear when the test scores and high school record are combined in the prediction of college performance. The median correlations for minority students exhibit the same pattern as that for white students. Except for minority males, the high school record tends to yield somewhat higher correlations than test scores. But the highest correlation is obtained when the test score and the high school record are combined to predict college performance. Unlike the situation for the white samples—where the advantage of adding the test score was about equal for all types of samples—the advantage for minority males (.15 points in the median correlation) is greater than for minority females (.07 points in the median correlation).

Minority/nonminority comparisons of median correlations

Combining the collective experience of the present survey, the Schrader (1971) results, and the Ford and Campos (1977) results, it is possible to estimate the relative predictive capabilities of the high school record and the test scores for nonminority males and females. These can then be compared with the figures developed above for minorities by combining black and Chicano samples. Considering that the Schrader medians are based on 51 colleges, the Ford and Campos medians on even larger numbers of colleges for 11 different years, and the nonminority median of the present study on 206 different samples, *averages* of these median correlations should be quite stable. Adding the minority medians from Table 7 and then comparing these with the nonminority medians suggests similar patterns of relationship for both. Consider the summary of the median correlations shown below:

	HSR		Tests and HSR	
	M	F	M	F
The present study (from Table 5)	.47	.51	.55	.56
Ford and Campos (1977)	.46	.52	.53	.60
Schrader (1971)	.47	.51	.55	.62
Average of nonminority median correlations	.47	.51	.54	.59
Minority median correlations (from Table 7)	.41	.57	.56	.64

For both minorities and nonminorities, the high school record would appear to yield higher correlations with college performance for females than for males. For both male and female minorities and nonminorities, the use of test scores increases the correlation with college performance.

Summary of correlational comparisons across studies

Median correlations were compared for all reported correlations for Anglo, black, and Chicano populations. Within each ethnic population, subclassifications by sex were made where possible. Because of limited data black and Chicano correlations were summarized as minority populations. Some possibly important sex differences were suggested within each of these subgroupings. Nonminority women tend to produce higher median correlations between predictor and criterion for all types of predictors examined. Minority women tend to yield higher correlations between all predictors and college performance than minority men. For all groupings except black males, the high school record yields higher correlations with college performance than test scores. But for all populations the combination of the high school record and test scores yields the highest median correlations with college perfor-

Table 7. Comparison of high school record and test scores as predictors of freshman or cumulative grade-point average for minority (or predominantly minority) male and female samples

Predictor	Males	Females	Combined sex samples	All samples
Number of samples				
High school record	23	21	36	80
Verbal test scores	13	12	31	56
Quantitative test scores	13	12	31	56
Composite test scores	9	9	28	46
HSR and tests	44	44	43	129
Range of correlations				
High school record	.11–.63	.25–.66	.02–.61	.11–.66
Verbal test scores	.23–.54	.17–.51	.06–.65	.06–.65
Quantitative test scores	.23–.46	.25–.49	.01–.50	.01–.50
Composite test scores	−.32–.54	.27–.64	−.04–.65	−.32–.65
HSR and tests	.23–.78	.29–.72	.11–.70	.11–.78
Median correlations				
High school record	.41	.57	.38	.46
Verbal test scores	.41	.42	.26	.36
Quantitative test scores	.40	.38	.18	.32
Composite test scores	.41	.52	.29	.34
HSR and tests	.56	.64	.42	.54

Note: See Appendix Tables A-3 and A-4 for a listing of the studies upon which this table is based.

mance. For all correlations reported for all samples, the following median correlational summary results (numbers of correlations used in computing medians are given in parentheses):

Predictor	Nonminorities	Minorities
High school record (HSR)	.48 (142)	.46 (80)
Verbal test scores	.38 (98)	.36 (56)
Quantitative test scores	.35 (104)	.32 (56)
HSR and tests	.54 (206)	.54 (129)

Correlational comparisons within studies

Another way of making correlational comparisons for different groups is to consider only those studies that have reported correlations based on data collected for two or more different groups. While this approach appears to have an advantage over the comparisons made in the previous section, it limits the number of samples that can be brought into the analysis. And while there exists the illusion of control, the samples being compared are not always comparable. Moreover, the sample sizes tend to be smaller for each group when two or more groups are considered in the same study. Nevertheless, a review of those studies that have included correlational comparisons of the various predictors for different groups is of interest for purposes of comparison with the across-studies comparisons.

Within-studies comparisons of the high school record as a predictor

In Table 8, 12 studies reporting correlational comparisons of the high school record as a predictor for Anglos and blacks, Anglos and Chicanos, or for Anglos, blacks, and Chicanos are summarized. Only two of the 12 studies collected data for all three groups (Goldman and Hewitt, 1976; Scott, 1976). Considering male and female comparisons separately, a total of 34 comparisons can be made from Table 8 of the high school as a predictor of college performance for minorities vs. nonminorities. In 21 of these comparisons the observed correlation for nonminorities is greater than the observed correlation for minorities, in 11 the opposite occurs, and two are even. The median correlation for Anglos and Chicanos is about the same, whereas that for blacks is substantially smaller. Therefore, the conclusion that is suggested from Table 8 is the same as that suggested from the analyses of correlations across studies: the high school record (used alone) appears to be a better predictor of college performance for nonminorities than it is for minorities. Note that the Chicano median of .36 is based on only eight correlations and is, therefore, not realistically comparable to the Anglo and black medians.

Within-studies comparisons of verbal test scores as predictors

Table 9 summarizes those studies that have reported correlations of verbal test scores and college performance for the groups of interest. The median correlation for blacks is highest, that for Anglos second highest, and that for Chicanos third. This comparison suggests that verbal test scores may be most useful for blacks and least useful for Chicanos. This suggestion is similar to that obtained from the across-studies comparisons in Tables 5 and 7, presented previously. The across-studies comparisons were based on many more samples of nonminorities and several more

Table 8. Studies comparing the high school record as a predictor of college performance for Anglo, black, and Chicano samples

Study[a]	Correlation coefficients								
	Anglo			Black			Chicano		
	M	F	T	M	F	T	M	F	T
Bowers (1970)	.35	.34		.14	.25				
Calkins and Whitworth (1974)	.42	.38					.36	.34	
Cleary (1968)			.66			.61			
Cleary (1968)			.64			.51			
Davis and Kerner-Hoeg (1971)									
Sample 1			.37			.22			
Sample 2			.43			.46			
Sample 3			.55			.22			
Sample 4			.51			.45			
Sample 5			.42			.06			
Sample 6			.33			.40			
Goldman and Hewitt (1975)									
Sample 1			.37						.37
Sample 2			.40						.42
Sample 3			.43						.21
Sample 4			.30						.36
Goldman and Hewitt (1976)			.36			.27			.37
Harris and Reitzel (1967)			.44			.50			
Pfeifer and Sedlacek (1971)	.63	.65		.23	.60				
Scott (1976)			.31			.36			.26
Wasserman (1978)			.21			.27			
Wilson (1978)[b]									
Sample 1			.31			.23			
Sample 2 (Soph cum)			.34			.19			
Sample 3 (Jnr cum)			.33			.26			
Sample 4 (Snr cum)			.32			.15			
Sample 5 (Sem 7)			.18			.07			
Sample 6 (Sem 8)			.20			.16			

Table 8 continued

Study[a]	Correlation coefficients: Anglo M	Anglo F	Anglo T	Black M	Black F	Black T	Chicano M	Chicano F	Chicano T
Wilson (1978)[b]									
Sample 7: FGPA			.53			.56			
Sample 8: cum GPA			.53			.59			
Sample 9: Sem 7 GPA			.35			.47			
Sample 10: Sem 8 GPA			.35			.35			
Median Correlations	.37			.27			.36		

a. Unless otherwise indicated, outcome predicted was freshman grades.

b. Compared minority and nonminority groups, but minority predominantly black.

samples of minorities and, consequently, it was possible to compare males and females somewhat more effectively. A previous suggestion from Tables 5 and 7, that minority males may be more predictable from verbal test scores than are nonminority males, is supported by the few comparisons of this type that can be made from Table 9. In six of the nine possible comparisons in Table 9 where correlations are reported for minority and nonminority males, the minority observed correlation is higher. But for females, the opposite occurs. In seven of the nine correlations comparing minority and nonminority females, the nonminority observed correlation is greater. In only one of the nine female comparisons is the minority observed correlation greater.

Within-studies comparisons of quantitative test scores as predictors

Table 10 lists those studies reporting comparisons of quantitative test scores as predictors for different ethnic groups. A total of 46 comparisons are possible and in 29 of these the observed nonminority correlation is greater, in 15 the minority correlation is greater, and in 2 the figures are the same. The median correlations are identical for Anglos and blacks, while the median correlations for Chicanos is somewhat lower. The across-studies comparisons of Tables 5 and 7 had indicated a slightly better overall correlation between quantitative test scores for nonminorities than for minorities (.35 vs. .32). As mentioned previously, this difference is possibly related to a proclivity among some minority students to avoid quantitative majors in their college work (see Altman and Holland, 1977).

Within-studies comparisons of composite test scores as predictors

Only five different studies were located that had comparisons of composite test scores as predictors of college performance for different ethnic groups. However, one study by Davis and Temp (1971) reported comparative correlations for 19 different samples. There are 28 possible comparisons between groups in the correlations of Table 11. In 22 of these comparisons the observed correlation is greater for nonminorities, suggesting that composite test scores considered alone—like either verbal or quantitative test scores considered alone, or high school record considered alone—do not correlate with college performance for minorities as well as they do for nonminorities.

Within-studies comparisons of the combination of HSR *and tests as predictors*

Table 12 shows summary results from 10 studies offering 63 different comparisons between minority and nonminority groups in which the high school record was combined with test scores to make predictions. When the high school record and test scores are combined, the differences between the groups tend to diminish—as was indicated earlier from the comparisons of Tables 5

Table 9. Studies comparing verbal test scores as predictors of college performance for Anglo, black, and Chicano samples

Study[a]	Test	Correlational coefficients								
		Anglo			Black			Chicano		
		M	F	T	M	F	T	M	F	T
Bowers (1970)	SCAT-V	.25	.28		.23	.17				
Calkins and Whitworth (1974)	SAT-V	.32	.37					.25	.34	
Cleary (1968)										
Sample 1	SAT-V			.45			.47			
Sample 2	SAT-V			.38			.26			
Sample 3	SAT-V			.47			.47			
Davis and Kerner-Hoeg (1971)										
Sample 1	SAT-V			.19			.13			
Sample 2	SAT-V			.35			.06			
Sample 3	SAT-V			.24			.26			
Sample 4	SAT-V			.31			.19			
Sample 5	SAT-V			.39			.43			
Sample 6	SAT-V			.29			.59			
Goldman and Hewitt (1975)										
Sample 1				.28						.16
Sample 2				.34						.42
Sample 3				.28						.25
Sample 4				.31						.22
Goldman and Hewitt (1976)	SAT-V			.30			.28			.16
Goldman and Richards (1974)										
Sample 1	SAT-V			.40						.33
Sample 2	SAT-V			.30						.24
Kallingal (1971)	CQT-V			.37			.43			
Pfeifer and Sedlacek (1971)	SAT-V	.51	.39		.56	.47				

and 7. In Table 12 the median of the 61 Anglo correlations is .48 and the median of the 38 black correlations is .47. The lower median correlation for Chicanos may be due to the smaller number of observations. The greater predictability suggested previously for nonminorities by individual measures, such as the high school record and different kinds of test scores, does not appear to be maintained when the high school record is combined with test scores in the multiple correlations. In Tables 5 and 7, where median correlations across studies were compared, the same kind of outcome was observed. The median correlation of all reported studies was the same for minorities and nonminorities (.54) when the high school record was combined with test scores. There does not appear to be any substantial difference, therefore, between minorities and nonminorities in the correlation between college performance and the combination of the high school record and test scores. *(Text continued on page 38.)*

Table 9 continued

Study[a]	Test		Correlation coefficients								
			Anglo			Black			Chicano		
			M	F	T	M	F	T	M	F	T
Stanley and Porter (1967)											
Sample 1	SAT-V		.42	.54		.38	.43				
Sample 2	SAT-V		.40	.55		.41	.32				
Sample 3	SAT-V		.35	.50		.42	.50				
Sample 4	SAT-V		.38	.47		.41	.32				
Sample 5	SAT-V		.31	.50		.41	.41				
Sample 6	SAT-V		.34	.48		.53	.40				
Wilson (1978)[b]											
Sample 1	SAT-V				.34			.24			
Sample 2	SAT-V	(Soph cum)			.38			.25			
Sample 3	SAT-V	(Jnr cum)			.41			.18			
Sample 4	SAT-V	(Snr cum)			.45			.24			
Sample 5	SAT-V	(Sem 7)			.32			.11			
Sample 6	SAT-V	(Sem 8)			.29			.09			
Sample 7	SAT-V				.41			.44			
Sample 8	SAT-V	(Cum GPA)			.42			.56			
Sample 9	SAT-V	(Sem 7 GPA)			.24			.57			
Sample 10	SAT-V	(Sem 8 GPA)			.32			.47			
Median correlations					.37			.41			.25

a. Unless otherwise indicated, outcome predicted was freshman grades.

b. Compared minority and nonminority groups, but minority predominantly black.

Table 10. Studies comparing quantitative test scores as predictors of college performance for Anglo, black, and Chicano samples

Study[a]	Test	Correlational coefficients								
		Anglo			Black			Chicano		
		M	F	T	M	F	T	M	F	T
Bowers (1970)	SCAT-Q	.31	.21		.14	.11				
Calkins and Whitworth (1974)	SAT-M	.33	.34					.23	.28	
Cleary (1968)										
Sample 1	SAT-M			.41			.01			
Sample 2	SAT-M			.30			.17			
Sample 3	SAT-M			.39			.47			
Davis and Kerner-Hoeg (1971)										
Sample 1	SAT-M			.21			.26			
Sample 2	SAT-M			.24			.30			
Sample 3	SAT-M			.15			.15			
Sample 4	SAT-M			.28			.09			
Sample 5	SAT-M			.37			.39			
Sample 6	SAT-M			.20			.40			
Goldman and Hewitt (1975)										
Sample 1				.21						.13
Sample 2				.26						.38
Sample 3				.24						.14
Sample 4				.22						.18
Goldman and Hewitt (1976)	SAT-M			.21			.18			.17
Goldman and Richards (1974)										
Sample 1	SAT-M			.37						.12
Sample 2	SAT-M			.25						.10
Kallingal (1971)	CQT-num.			.31			.20			
Pfeifer and Sedlacek (1971)	SAT-M	.41	.35		.46	.49				

Table 10 continued

Study[a]	Test		Correlational coefficients: Anglo M	Anglo F	Anglo T	Black M	Black F	Black T	Chicano M	Chicano F	Chicano T
Stanley and Porter (1967)											
Sample 1		SAT-M	.47	.55		.34	.43				
Sample 2		SAT-M	.39	.55		.25	.26				
Sample 3		SAT-M	.38	.54		.40	.41				
Sample 4		SAT-M	.40	.49		.41	.36				
Sample 5		SAT-M	33	.54		.33	.43				
Sample 6		SAT-M	.45	.45		.40	.36				
Wilson (1978)[b]											
Sample 1	SAT-M				.30			.19			
Sample 2	SAT-M	(Soph cum)			.32			.14			
Sample 3	SAT-M	(Jnr cum)			.32			.13			
Sample 4	SAT-M	(Snr cum)			.33			.18			
Sample 5	SAT-M	(Sem 7)			.22			.03			
Sample 6	SAT-M	(Sem 8)			.13			.01			
Sample 7	SAT-M				.40			.44			
Sample 8	SAT-M	Cum GPA			.40			.55			
Sample 9	SAT-M	Sem 7 GPA			.24			.57			
Sample 10	SAT-M	Sem 8 GPA			.32			.47			
Median correlations				.33			.33			.17	

a. Unless otherwise indicated, outcome predicted was freshman grades.

b. Compared minority and nonminority groups, but minority predominantly black.

Table 11. Studies comparing composite test scores as predictors of college performance for Anglo, black, and Chicano samples

Study[a]	Test	Correlation coefficients[a]								
		Anglo			Black			Chicano		
		M	F	T	M	F	T	M	F	T
Calkins and Whitworth (1974)	SAT	.38	.40					.26	.35	
Davis and Temp (1971)										
Sample 1	SAT			.27			.26			
Sample 2	SAT			.23			.15			
Sample 3	SAT			.38			.30			
Sample 4	SAT			.33			.18			
Sample 5	SAT			.55			.51			
Sample 6	SAT			.39			.25			
Sample 7	SAT			.15			.07			
Sample 8	SAT			.51			.27			
Sample 9	SAT			.45			.06			
Sample 10	SAT			.25			.39			
Sample 11	SAT			.43			.08			
Sample 12	SAT			.49			.41			
Sample 13	SAT			.46			.15			
Sample 14	SAT			.25			.26			
Sample 15	SAT			.38			.31			
Sample 16	SAT			.27			.28			
Sample 17	SAT			.34			.19			
Sample 18	SAT			.45			.48			
Sample 19	SAT			.31			.65			
Goldman and Richards (1974)										
Sample 1	SAT			.44			(.44)[b]			
Sample 2	SAT			.33			.24			
Kallingal (1971)	CQT			.53			.51			
Pandey (1971)	SCAT	.12	.66		−.32	.27				
Scott (1976)	ACT			.21			−.04			.20
Median correlations			.38			.26			.26	

a. Outcome predicted was freshman grades.

b. This figure was not reported but it was stated that "the difference between the (multiple) correlation coefficients was nonsignificant."

Table 12. Studies comparing multiple correlations of high school record combined with test scores as predictors of college performance for Anglo, black, and Chicano samples

Study[a]	Correlation coefficients								
	Anglo			Black			Chicano		
	M	F	T	M	F	T	M	F	T
Bowers (1970)	.41	.39		.30	.29				
Calkins and Whitworth (1974)	.47	.46					.38	.41	
Davis and Kerner-Hoeg (1971)									
Sample 1			.41			.31			
Sample 2			.53			.50			
Sample 3			.55			.36			
Sample 4			.54			.46			
Sample 5			.56			.49			
Sample 6			.43			.70			
Farver, Sedlacek, and Brooks (1975)									
Sample 1	.63	.65		.61	.67				
Sample 2 (Soph GPA)	.61	.73		.39	.63				
Sample 3 (Soph cum)	.68	.76		.51	.71				
Sample 4 (Jnr GPA)	.45	.50		.47	.36				
Sample 5 (Jnr cum)	.57	.73		.52	.61				
Sample 6 (Snr GPA)	.28	.39		.40	.35				
Sample 7 (FGPA)	.61	.55		.44	.57				
Sample 8 (Soph GPA)	.50	.42		.31	.50				
Sample 9 (Soph cum)	.59	.49		.47	.59				
Sample 10 (Jnr GPA)	.53	.36		.23	.43				
Sample 11 (Jnr cum)	.64	.50		.78	.61				
Goldman and Hewitt (1975)									
Sample 1			.43						.38
Sample 2			.48						.49
Sample 3			.47						.28
Sample 4			.41						.37
Goldman and Hewitt (1976)			.43			.33			.38
Scott (1976)			.32			.38			.29
Silverman, Barton, and Lyon (1976)									
Sample 1			.44			.31			
Sample 2			.39			.49			

Table 12 continued

Study[a]	Correlation coefficients: Anglo M	Anglo F	Anglo T	Black M	Black F	Black T	Chicano M	Chicano F	Chicano T
Warren (1976)									
Sample 1	.13	.49	.32				.58	.56	.55
Sample 2	.29	.49	.41				.35	.51	.43
Sample 3	.43	.28	.32				.13	.37	.31
Sample 4	.56	.35	.49				.58	.46	.48
Sample 5 (Jnr GPA)			.32						.29
Sample 6 (Jnr GPA)			.31						.32
Sample 7 (Snr GPA)			.34						.52
Sample 8 (Snr GPA)			.33						.11
Sample 9 (UD[b] GPA)			.49						.41
Wilson (1978)[c]									
Sample 1			.51			.41			
Sample 2 (Snr cum)			.54			.42			
Sample 3			.57			.60			
Sample 4 (Snr cum)			.57			.69			
Median correlations		.48			.47			.38	

a. Unless otherwise indicated, outcome predicted was freshman grades.

b. Upper division.

c. Compared minority and nonminority groups, but minority predominantly black.

Summary of correlational comparisons within studies

When comparisons of correlations are limited only to those reported within a given study, the number of correlations available is somewhat less than was the case of the comparisons discussed in the previous section. Nevertheless, the summarizations yield similar results. With the smaller number of correlations to work with, few generalizations about sex groups were possible. As before, the high school record tended to yield higher median correlations with college performance for nonminorities than for minorities, and this was the case also for verbal test scores and quantitative test scores. These median comparisons are summarized below, with numbers of correlations given in parentheses (note, however, that the medians for Chicanos are based on small numbers of samples):

	Anglo	Black	Chicano
High school record (HSR)	.37 (32)	.27 (26)	.36 (8)
Verbal test scores	.37 (45)	.41 (37)	.25 (9)
Quantitative test scores	.33 (45)	.33 (37)	.17 (9)
HSR and tests	.48 (61)	.47 (38)	.38 (25)

These comparisons suggest, as have previous comparisons of this report, that differences between the predictability of the groups of interest are minimized when the high school record and test scores are combined for predictive purposes.

Special issues in prediction

The preceding sections have touched on, at least implicitly, a number of special issues in prediction that are perhaps worthy of further elaboration. Some of these have been extensively discussed in the literature while others have not, but all arise out of issues that are basic in considerations of population validity. Most of these special issues have not been adequately researched and, accordingly, it is not possible to draw many firm conclusions.

Are predictive validities declining?

It has been suggested that predictive correlations may have declined in the past several years as a result of test score declines, grade inflation, societal changes, and related factors. Ford and Campos (1977) addressed this issue and found no change in predictive relationships for the years studied (1964 through 1974). Results of the Ford and Campos studies are tabulated as a part of Appendix Table A-1. Ford and Campos studied males and females, but did not study ethnic populations. There is a question, therefore, as to whether a decline in predictive relationships may have occurred for minorities. Although no analysis comparable to that of Ford and Campos is available, some indication of the trend in predictive relationships for minorities may be obtained by a consideration of the published correlations given in Appendix Tables A-3 and A-4 for a recent time period. Given the need to subdivide the samples, and thus reduce the number available for analysis, only those studies reporting predictive correlations for the high school record combined with the test are considered. When all samples were combined, Table 6 shows that a total of 110 correlations were reported for black or predominantly black samples and a median correlation of .57 was computed from the total of 110. For studies with publication dates of 1970 or later, a total of 48 correlations of this type are reported. These later 48 correlations have a median of .48. Although crude, this kind of analysis suggests that there may have been a decline in predictive relationships for blacks even though there was no decline generally. Or it may be that the published correlations are not representative of national trends and are not comparable over time. No similar analysis can be conducted for Chicanos because all available studies are recent.

Long-range criteria

A second special issue of interest concerns long-range criteria, since it has been hypothesized that minority populations may be less predictable for the long-term than for short-term. More specifically, the issue is whether junior or senior college performance for minority populations is as predictable as freshman college performance. Some of the more recent studies have focused on this issue (Farver, Sedlacek, and Brooks, 1975; War-

ren, 1976; Wilson, 1978). One concludes from these studies that, while predictive relationships tend to be reduced for all populations as longer-range criteria are used, no systematic differences occur across groups. The reduced relationships may be the result of restriction in the range of the criterion measures, reduced reliability in the criterion measures, attrition of students, and various extraneous factors that enter the situation as time passes.

Individual versus group considerations

A third special issue relates to the use of measures and the impacts they have on different groups of individuals. It has been shown by Thorndike (1971) that uses of measures that are fair to individuals may not be fair to groups of individuals. Some reanalyses of data presented previously on over- and underprediction has suggested that the traditional uses of measures in

Table 13. Prediction studies using nontraditional methods

Study	Models or analyses	Summary
Breland (1977)	False negatives Hits and misses Objective criterion	False negative proportions and hits and misses compared for males and females and majority and minority students. No significant differences in false negatives. Misses generally less for minorities using both criteria.
Breland and Ironson (1976)	Constant ratio Conditional probability Equal risk	Analysis of DeFunis data shows that all models would admit substantially fewer minorities than actually admitted.
Cole (1973)	Constant ratio Conditional probability (CP) Equal risk (ER) Subjective regression Traditional regression (TR)	Analysis of hypothetical data shows CP model most favorable to minorities and ER and TR least favorable.
Goldman and Widawski (1976)	False negatives	False negative proportions computed for Blacks and Chicanos using HSR alone and HSR and SAT in combination. Suggests that SAT increases proportions of Black and Chicano false negatives.
Linn (1973)	Constant ratio	Reanalysis of Davis and Kerner-Hoeg (1971) and Temp (1971) data using Thorndike model suggests bias against blacks taken as a group.
Schmidt and Hunter (1974)	Constant ratio	Reanalysis of Cleary (1968) and Temp (1971) data suggests bias against blacks taken as a group.
Silverman, Barton, and Lyon (1976)	Constant ratio	No bias against blacks when Thorndike model used with a common regression line.

selection may be unfair to low-scoring groups (Linn, 1973; Schmidt and Hunter, 1974). Thorndike proposed a model for selection that would avoid unfairness to low-scoring groups by the use of a constant ratio of proportion selected to proportion successful for all relevant groups. However, the Thorndike constant ratio model has been questioned on the grounds that it would not maximize overall performance (Schmidt and Hunter, 1974), and that it is mathematically indefensible (Petersen and Novick, 1976). Other selection models have also been proposed by Cole (1973), and Darlington (1971), but these models have also been seriously questioned.

One way of gauging the significance of nontraditional approaches to the analysis of selection measures, such as the new selection models proposed, is to consider which approaches have been applied to data and reported. Table 13 summarizes several studies that have explored nontraditional methods. As indicated, five of the papers summarized have applied the Thorndike constant ratio model to data, but by and large there has not been much empirical research using these models. Some interest is also indicated in the analysis of false negatives (persons rejected but who could have succeeded) in the research by Breland (1977) and Goldman and Widawski (1976). The lack of research may be a result of the difficulty in conducting it, or it may be a result of a failure to view the theoretical models seriously.

Criterion bias

The issue of criterion bias may be more important than evidenced by the published research. The one study indicated in Table 13 exploring this issue (Breland, 1977) attempted to minimize criterion bias by the use of a blind-scored essay examination. That is, the examinees and the scorers were unknown to each other and there was no identification of essays to indicate population membership. The similarity of the results to those of other studies (minorities overpredicted) suggested that criterion bias may be less of a problem than it was previously thought to be. Nevertheless, a single study can never be conclusive and, therefore, the issue appears to call for further research using objective criteria.

Internal analyses

The previous section viewed population validity in the context of an external criterion, that is, college performance. It is also possible to think of population validity in the absence of an external criterion. Internal analyses focus on the elements of a measure, clusters of elements, and the relationships between these elements and clusters and the meaning of total scores derived from a measure. There has been little analysis of this type for grades as a measure and, therefore, what follows in this section pertains only to tests as measures of college abilities. In considering the elements of a test (test items and subscores) in relation to population validity, it is important to recall that what is of principal interest in this report is the validity of inferences that are made from test scores. When one examines the items of a test it should be recognized that what is being studied are microscopic parts of a measure from which inferences are not normally made (because the item scores are not reported). Consequently, population validity as examined through internal analyses is most useful when reported portions of a measure are examined in their entirety. However, since there has been considerable attention given to item analysis procedures, for test development purposes, and since these procedures have implications for population validity, studies of this type are covered briefly in this section. No attempt is made to review thoroughly item analysis procedures and studies of them. Readers interested in further comparative details of these procedures and studies using them should consult the review papers of Potthoff (1972), Angoff (1975, 1976), Hunter (1975), Rudner (1977), Petersen (1977), Subkoviak and Baker (1978), Lord (1977), Ironson (1978), Rudner and Convey (1978), and Merz and Rudner (1978). Other papers relating to particular procedures and applications of them are indicated in the following subsections.

Comparison of proportions responding correctly to items

This approach is one of the most well-known and widely used, perhaps because of its simplicity. Item data are collected for samples from two populations, difficulty indices computed, and a cross-plot then made on a Cartesian plane. Item difficulties are sometimes retained in the form of p-values (proportion of a group answering an item correctly) for the cross-plots (Green and Draper, 1972). Angoff and Ford (1973), however, advocate an inverse normal transformation of the p-values to normal deviates prior to plotting the difficulties on a bivariate graph. Lord (1976) notes that this transformation tends to remove the curvature in the relationship between the item difficulties for the two groups, but that it introduces errors as a result of different guessing patterns in the groups being compared. Regression lines may be com-

puted for the best fit for each group with the associated correlation coefficient representing an index of the degree to which item difficulties are similar for the two groups being compared.

Research using this method is exemplified by the work of Angoff (1972), Angoff and Ford (1973), Green and Draper (1972), Breland et al. (1974), and Hicks et al. (1976). Angoff and Ford (1973) computed the correlations between PSAT transformed item difficulties for black and white groups and obtained correlations in the low .90s, but higher correlations were obtained when groups were matched on ability. Green and Draper (1972) showed that if biserial correlations were used as a covariate the correlation is increased. Subkoviak and Baker (1978) also noted that item-group interactions (proportional to correlations) decrease noticeably when comparison groups are matched on ability. Breland et al. (1974) cross-plotted transformed item difficulties for several groups tested in the National Longitudinal Study and obtained results similar to those of Angoff and Ford (1973). Beyond the empirical evidence, there are theoretical reasons (Hunter, 1975; Lord, 1976) to indicate that ability differences between groups may affect outcomes more than do other (e.g., cultural) differences. Because of its simplicity the method has been used as a means for initial screening of items in the test development stage, but there is always a question regarding the degree of aberrancy that is important. Both Hunter (1975) and Lord (1976) have argued that the comparison of proportions responding correctly in different groups (with or without the inverse normal transformation) leads to false inferences because of statistical artifacts in the method. In Hunter's (1975) view, the method is useful only if all items are equally discriminating for the groups being compared, there is no guessing, the items are unidimensional, the groups compared are of equal ability, and the ICC (see subsection "the item characteristic curve method," below) curves are normal ogives.

The item characteristic curve method

This approach is derived from latent trait theory and involves the comparison of item characteristic curves (ICCs) for different groups. The item characteristic curve results from a plot of true ability, achievement, or total score (on the abscissa) versus the probability of a correct item response (on the ordinate). The probability of a correct response increases monotonically (usually) as ability increases, and the classical curve has an inflection point. The slope of the curve at the inflection point indicates the power of the item to discriminate among persons of similar ability levels. The position of the inflection point along the abscissa indicates the degree to which ability is necessary for a correct response. The ICC method can be used to compare the curves for samples from different populations. Ideally, all populations should have equal probabilities of responding correctly, given the same true ability level. The ICC approach is appealing to statistical sophisticates, but there are practical problems, primarily economic, with using it as a routine item screening procedure. Moreover, as with the previously described item analysis methods, there are conceptual problems with using the ICC method within the framework of population validity. Lord (1977) plotted ICC curves for blacks and whites for SAT-V items and obtained some puzzling results. A total of 38 of the 85 SAT-V items analyzed had statistically different (.05) ICC curves, but an examination of some of the most "different" items in detail provided no insight as to why the ICC curves were different. These results, as well as results from the previously described item analysis methods, reveal little about population validity. Since inferences are made from combinations of items (reported subscores and total scores), it is the combination of items, not individual items, that internal analysis must focus on for purposes of population validation.

The analysis of variance method

The procedure is to sample two or more populations and then to examine the variance that can be attributed to various sources. Four principal sources of variance are of interest (Rudner, 1977). First, variance due to differences in items is considered because some items are more difficult than others. Second, variance due to group differences occurs because groups may not be equivalent with respect to the measured attribute. Third, variance due to ability differences of subjects within groups is examined. Fourth, and most important, variance resulting from an interaction of items and groups is isolated. Ideally, the item-group interaction variance will be small. Cardall and Coffman (1964) used a two-factor analysis of variance design to compare Southeastern black students, Midwestern rural students, and New York City students on the SAT. Statistically significant item-group interactions were obtained for both SAT-V and SAT-M items. Cleary and Hilton (1968) made comparisons of black and white students on the PSAT and also found significant item-group interactions, but they also found that percentage of variance accounted for by the item-group interaction amounted to less than 2 percent of the total variance. They concluded, therefore, that the PSAT item-group interaction was of no practical significance. Other studies using the analysis of variance method include Eagle and Harris (1969), Hoepfner and Strickland (1972), Green and Draper (1972), and Jensen (1973). The analysis of variance method has some conceptual limitations. One of these limitations is that the item-group interaction may be more sensitive to ability differences than it is to other group differences. Rudner (1977) notes also that the method may have a tendency to classify erroneously individual items as biased.

The chi-square method

Scheuneman (1975, 1976) and Green and Draper (1972) have examined this approach. This method attempts to determine whether people of the same ability level have the same probability for a correct response independent of group membership. Samples are divided into groups on the basis of observed scores and the proportions within each level responding correctly are compared by means of a chi-square test. The proportion of correct responses for each group is expected to be the same for all persons in the same total score interval. Because the procedure requires a large number of inference tests, incorrect inferences may be made because of inflated significance levels. Rudner (1977) and Hunter (1975) note as well that the method appears to be unduly sensitive to population differences in total score distributions.

The distractor response analysis method

All of the methods described above assume that item responses are scored either correct or incorrect, but it is possible to analyze all responses. This approach has been recommended by Veale and Foreman (1976) as a means of examining items for different population groups. In the event that a single group is overly attracted to a particular distractor, there may exist some undesirable characteristic of the item. Comparing distractor response rates for different groups by a chi-square test gives an indication of cross-group similarity, or dissimilarity in responses. Veale and Foreman (1976) suggest other kinds of statistics as well. The distractor response analysis approach assumes that any undesirable characteristics of an item are caused by distractors. It ignores the possibility that the item stem, the directions, or the subject matter may be the source of undesirable item characteristics. And like some of the other item analysis methods, distractor response analysis uses large numbers of inferential tests and, accordingly, the probability of inferential

errors is relatively high. On the other hand, distractor response analysis is the only one of the item analysis methods available that is suitable for criterion-referenced tests (Rudner, 1977).

Factor analysis methods

Factor analytic approaches have been suggested both for item screening and for construct validation. For routine screening, economic problems (as in the ICC method) are a limitation. Ignoring such limitations, items that have large differences across populations in factor loadings can be eliminated during test development. The factor loadings are item-trait correlations and, ideally, similar correlation matrices will result from factor analyses conducted on samples from different populations. Rudner (1977) notes, however, that factor analytic approaches to item analysis are not without some conceptual and practical limitations. To obtain stable correlation matrices for use in factor analysis large numbers of subjects are needed. There are also questions as to which type of correlation to use. When dichotomously scored variables are involved, either phi or tetrachoric correlations result and each has limitations. The phi coefficient is limited because it is highly sensitive to item difficulties and tetrachoric correlations are known for their instability. Moreover, numerous decisions are required in factor analysis (which method, what rotation, how many factors) and different results are obtained as a result of these decisions. Thus, like the item characteristic curve approach, the factor analytic approach to item analysis is appealing but of limited practical utility. Hunter (1975) argues that factor analytic methods are inappropriate for item analysis because of the nonlinearity of item characteristic curves.

Fewer criticisms are made when factor analytic methods are used to examine factorial consistency of entire tests or other entire measures across populations. Hennessy and Merrifield (1976) factor analyzed the Comparative Guidance and Placement (CGP) Tests separately for samples of black, Spanish-American, Jewish, and Caucasian populations. Highly similar factor solutions were obtained for each group, and although some differences in the magnitude of factor loadings were found, the basic structures were almost identical. They concluded that the cross-ethnic construct validity of the CGP and the relative invariance of ability structures across groups was demonstrated. Similar results have been obtained using factor analysis on other measures by Mitchell (1956), Semler and Iscoe (1966), and Hollmann (1973).

Promising research along factor analytic lines has recently been completed by Rock and Werts (1979). This type of analysis is unique, however, in that it provides statistical tests for equality of factor patterns and equality of units of measurement for different population groups. Rock and Werts examined SAT-V and SAT-M scores for American Indians, blacks, Mexican Americans, Oriental Americans, Puerto Ricans, and whites. They concluded that the four item types in the SAT-V (antonyms, reading, sentence completion, and analogies) and the two item types in the SAT-M (regular math and quantitative comparisons) appear to measure all the populations studied with equal accuracy.

Analysis of content

Ebel (1975) has suggested that population validity is derived to a large degree from content validity and, furthermore, that for achievement tests the sole criterion for validity is ultimately content validity. From this perspective group differences resulting from learning opportunities, language usage, and other factors are considered as explanations for performance differences, but they are not evidence of population invalidity. To insure content validity one must:

Clearly define a domain of tasks.

Give the test a name that indicates the principal features of that domain.

Sample representatively from that domain.

Demonstrate the reliability of the scores on the test from a representative sample of the population of intended examinees.

Thus the content validity of a test is largely determined, in this view, in the test development stage. If a test is developed with care and labelled properly, then there can be no question of population validity, according to Ebel, as long as it is used as an achievement test. Hunter (1976) takes issue with the Ebel position and argues that, in addition to careful test development, empirical studies of test content are necessary.

Willingham (1974) and others have emphasized the need for expert judgment, with data collected objectively, for content validation. There exist few reported studies in the literature, however, offering objective evidence of content validity. Breland, Conlan, and Rogosa (1976) report a study of the Test of Standard Written English (TSWE) in which English professors rated the importance of specific types of content (some included in the TSWE and some not included) for freshman English placement. The results showed that the types of content contained in the TSWE tend to agree with the content professors consider to be the most important. A content validation study was also recently conducted for the National Teacher Examinations (NTE), and the content validity of the test was confirmed (Educational Testing Service, 1976). Bianchini (1976) reported a study in which minority judges suggested which items should be removed from certain achievement tests. The interjudge agreement on which items should be removed was quite low and indicated, therefore, that the elimination of items in this way would have little systematic effect on the total test score. Other unpublished studies using a panel of judges to evaluate items for inappropriateness for some populations have yielded similar results. Flaugher and Schrader (1978) removed a substantial number of items from the SAT-V and SAT-M that were appreciably more difficult for black than for white students and found that removal of the items had little effect on total score differences between black and white groups. In fact, removing these items made the tests more difficult for both groups.

The limited amount of evidence published on content validation suggests that more objective studies should be undertaken. There is also a need for more descriptive information concerning the test development process, the history of the development of a test, and the philosophy and rationale underlying a testing program, such as that contained in Angoff (1971) for College Board tests.

Summary of internal analysis methods

Most of the methods discussed in this section concentrate on individual test items, and it is from that limited focus that most of their difficulties arise. From the perspective of population validity, however, the appropriate level of analysis is the composite of individual components (e.g., items) of a measure and not the item level. This follows from the fact that inferences are not made as a result of performances on individual items but from performances on tests, or other measures, as a whole. Actually, inferences are usually made in the college entrance context from more than one measure, such as grades and test scores in combination. The item analysis approaches are primarily of interest in the test development stage as a means of routinely screening items for optimum characteristics. But Ironson (1978) has demonstrated that different procedures identify different items for deletion. For the study of population validity, confirmatory factor analysis of reported total scores and subscores for all groups of interest is seen as one promising approach. Another approach to internal analysis suggested is that of objective content analysis based on the informed judgment of experts in a field.

Summary and conclusions

In this report an attempt has been made to review the population validity of commonly used college entrance measures as represented by reports and published articles. The large number of relevant documents makes the task difficult, but it is believed that a useful compilation of information has resulted. By viewing the problem from the perspective of population validity (rather than the more common test bias approach) a broader and less equivocal set of statements can be made about what is known. Examined are not only tests, as traditionally conceived, but also the high school record—since seldom is either a test score or the high school record used alone in making inferences about probable college performance. The presentation of results from empirical validation studies is preceded by a review of some of the problems encountered in interpreting validity study results. Attention is given to the problem of developing or locating adequate criteria for use in validation studies, and particular note is made of the problems specific to the college grade-point average. Other interpretational problems such as those caused by restriction of range and pooled data samples are also described.

Two basic approaches were used to examine the predictive validity of college entrance measures for different populations: regression analysis and correlational analysis. In the regression approach, the principal interest is with the over- or underprediction of performance for various groups based on data either from another group or from a combination of groups that are different from the population being examined. The correlational analyses compare magnitudes of predictive correlations observed for samples of different populations.

The summaries of regression analyses indicated a fairly consistent overprediction of college performance for black samples when the regression equation was developed either from white samples or predominantly white samples. For women, however, an opposite kind of error in prediction tends to occur consistently. If a regression equation developed from data on male or predominantly male samples is used to predict college performance for women, underpredictions tend to occur. The regression analyses for Chicanos are less conclusive because of the limited numbers of studies that have been published, but overpredictions similar to those for black samples are indicated.

The high school record, when used alone, appears to result in the largest overpredictions, even though tests used alone also result in overpredictions for minority group members. These overpredictions occur both for short-range criteria (freshman college performance) and long-range criteria (senior college performance), but the amount of data available for the longer-range predictions are limited. Overpredictions are mini-

mized when the high school record is used in combination with test scores.

Two kinds of correlational summaries of reported data were made. In the first, median correlations were computed across all available studies for the different populations of interest. That is, the median correlational summaries were not limited to studies in which two or more groups were compared but included all studies where a population of interest was studied. In the second kind of correlational analysis, the summarization was limited to studies comparing population samples within institutions. The median correlations for the latter type of comparison were less stable, but tended to agree with the former kind of comparisons.

The across-studies median correlational comparisons showed, for white samples, a tendency for the high school record to provide slightly higher correlations with college performance than test scores. They also indicate a tendency for white female samples to yield higher predictive correlations than white male samples. For white males or white females or for combined sex groups, the median predictive correlations based on a combination of the high school record and test scores are higher than are those based on single predictors. For black samples, the median correlational comparisons also resulted in slightly higher median predictive correlations for the high school record as compared to test scores. Within black sex groups, however, the higher medians for the high school record occurred only for females. Both black males and black females yielded higher median predictive correlations when the high school record was combined with test scores.

In the within-studies correlational comparisons, Anglo populations were contrasted with black and Chicano populations. For this group of studies, the high school record had the lowest predictive correlation median for black populations. Verbal test scores had the highest median predictive correlation for black populations. Quantitative test scores had the lowest predictive correlations for Chicanos, with blacks and Anglos having equivalent median correlations. When the high school record was combined with test scores, however, the median Anglo and median black correlations were essentially the same.

A review of several special topics in prediction considered issues of possibly declining validities, long-range criteria, individual vs. group outcomes, and criterion bias. Although some of these issues have been discussed at length in the literature, evidence seems lacking with respect to most. There is some evidence suggesting that predictive validities have not declined in recent years for combined populations and that long-range criteria operate similarly for majority and minority populations. These questions suggest a need for further investigation, however.

Several methods of internal analysis of tests were briefly described, criticisms of these methods reviewed, and results using the methods were surveyed. Considerable difficulty with most of these methods was noted, and it was concluded that those methods concentrating on individual items are more relevant to the optimization of item characteristics than to population validity. This conclusion followed from the accepted notion that the primary concern should be with the validity of inferences made from the results of measures. From this perspective one does not validate tests or test items but inferences made as a result of reported data based on the tests or items. Since item scores are not usually reported, inferences are not usually made from them. Inferences are only made from reported total scores and subscores, and more often than not, inferences are made from those only in the context of still other information such as previous academic record. The more useful approaches to internal analysis were seen to be analyses of factorial consistency across populations, objective studies of content based on expert judgment, and philosophical and historical analyses of the ra-

tionale of specific content. Several studies reporting factorial consistency evidence of population validity for college entrance (or similar) tests were cited. Two studies supporting population validity from an expert judgment perspective were cited. Further work on both factorial consistency and expert judgment, as well as on the content rationale of college entrance measures, was suggested.

Some specific kinds of research projects have been suggested by this review, and it is useful to list and summarize these briefly.

1. *Long-range prediction.* Several studies cited in this review suggested that population validity was maintained when the criterion data were collected after a longer interval. Senior GPA, for example, was predicted about as well for all populations studied. Further studies of this type would be useful.

2. *Major field studies.* The proclivity for members of different populations to choose different major fields systematically suggests that prediction studies limited to certain major fields or course groupings would yield some interesting findings.

3. *Studies of the college record.* The general limitations of the college GPA as a criterion and possible differences in the meaning of college GPA for different groups suggests that efforts are needed to investigate the characteristics of present criteria.

4. *Studies of the high school record.* Most past research has concentrated on the predictive capabilities and the reliability of college entrance tests and has tended to ignore other measures, such as the high school record, commonly used at the time of college entrance. Some of the evidence of this review would indicate differential population validity (and perhaps reliability) of the high school record.

5. *Neglected populations.* Although a few studies have been made of Hispanic groups, these are not sufficient to allow for any sound generalizations. Hispanics are often grouped with blacks to construct a minority population. Given the possibility of important linguistic influences, it seems essential that more studies be made—both in prediction and in internal analysis—for groups having had substantially different linguistic and cultural experiences. A more serious problem of neglect exists with respect to handicapped people and for older populations.

6. *Item characteristic curves.* Notwithstanding the limitations noted in this review of item characteristic curve theory as a source of population validation, the theoretical advantages of the method for the optimization of item characteristics suggest further research. In particular what appears to be needed is research comparing not only sex and ethnic groups, but research comparing SES, geographical, linguistic, and other groups within the same overall samples used for the sex and ethnic comparisons.

7. *Factorial consistency of measures.* Both predictor and criterion measures need to be examined with respect to their factorial consistency across significant populations.

8. *Expert judgment of test content.* Carefully designed objective studies of test content can be designed to elicit expert judgment concerning the content of college entrance measures. For some kinds of measures, such as achievement tests, expert judgment is the most useful type of validation information.

9. *Content rationale of tests.* The content of different tests used varies considerably and, therefore, there are most likely different rationales available to support each. Probably, much of this rationale is contained in the archives of psychological and educational literature, but it should be retrieved, summarized, and made available to the users of tests.

References

Altman, R. A., and Holland, P. W. *A summary of data collected from Graduate Record Examinations test-takers during 1975–76.* GRE Data Summary Report No. 1. Princeton, N.J.: Educational Testing Service, 1977.

American College Testing Program. Bias in the uses of educational assessment. In *Assessing Students on the Way to College* (Vol. 1). Iowa City, Iowa: Author, 1973.

American Psychological Association, Inc. *Standards for educational and psychological tests.* Washington, D.C.: Author, 1974.

Anastasi, A. *Psychological testing: Fourth edition.* New York: Macmillan Publishing Co., Inc., 1976.

Angoff, W. H. *The College Board Admissions Testing Program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests.* New York: College Entrance Examination Board, 1971.

Angoff, W. H. *A technique for the investigation of cultural differences.* Paper presented at a meeting of the American Psychological Association, September 1972.

Angoff, W. H. *The investigation of test bias in the absence of an outside criterion.* For presentation at NIE Conference on Test Bias, December 4, 1975.

Angoff, W. H. *General considerations in the study of test bias when the criterion is absent.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1976.

Angoff, W. H., and Ford, S. F. Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement,* 1973, *10*(2), 95–106.

Astin, A. W. *Predicting academic performance in college.* New York: The Free Press, 1971.

Bechtoldt, H. P. Construct validity: A critique. *American Psychologist,* 1959, *14*(10), 619–629.

Biaggio, A. M. B. *Relative predictability of freshman grade-point averages from* SAT *scores in Negro and white southern colleges.* Report No. BR-5-0216-TR-7. Research and Development Center for Learning and Re-education. Madison, Wis.: University of Wisconsin, 1966.

Bianchini, J. C. *Achievement tests and differentiated norms.* Paper presented at the USOE Invitational Conference on Achievement Testing of Disadvantaged and Minority Students for Educational Program Evaluation, Reston, Va., May 27–30, 1976.

Borgen, F. H. Differential expectations? Predicting grades for black students in five types of colleges. *Measurement and Evaluation in Guidance,* 1972, *4*(4), 206–212.

Bowers, J. The comparison of GPA regression equations for regularly admitted and disadvantaged freshmen at the University of Illinois. *Journal of Educational Measurement,* 1970, *7*(4), 219–225.

Boyd, W. M., II. SATs and minorities: The dangers of underprediction. *Change,* November 1977, pp. 48–49.

Bracht, G. H., and Glass, G. V. The external validity of experiments. *American Educational Research Journal,* 1968, *5*(4), 437–474.

Breland, H. M. *Group comparisons for the Test of Standard Written English.* College Board Research and Development Report (RDR-77-78, No. 1); also Research Bulletin (RB-77-15). Princeton, N.J.: Educational Testing Service, 1977.

Breland, H. M., Conlan, G. C., and Rogosa, D. *A preliminary study of the Test of Standard Written English.* Princeton, N.J.: Educational Testing Service, 1976.

Breland, H. M., and Ironson, G. H. *Defunis* reconsidered: A comparative analysis of alternative admission strategies. *Journal of Educational Measurement,* 1976 *13*(1), 89–98.

Breland, H. M., Stocking, M., Pinchak, B. M., and Abrams, N. *The cross-cultural stability of mental test items.* Project Report 74-2. Princeton, N.J.: Educational Testing Service, 1974.

Brown, J. L., and Lightsey, R. Differential predictive validity of SAT scores for freshman college English. *Educational and Psychological Measurement,* 1970, *30*(4), 961–965.

Calkins, D. S., and Whitworth, R. *Differential prediction of*

freshman grade point average for sex and two ethnic classifications at a southwestern university. El Paso, Texas: University of Texas (available through ERIC #ED 102–199), 1974.

Cardall, C., and Coffman, W. E. *A method for comparing the performance of different groups on the items in a test.* College Board Research and Development Report (RDR-64-5, No. 9); also Research Bulletin (RB-64-63). Princeton, N.J.: Educational Testing Service, 1964.

Chansky, N. M. Aptitude, personality, and achievement in six college curricula. *Educational and Psychological Measurement*, 1965, *25*(4), 1117–1124.

Chase, C. I., and Thompson, R. *Predicting success for university freshmen* (Indiana Studies in Prediction, No. 1, Supplement IV, Monograph of the Bureau of Educational Studies and Testing). Bloomington, Ind.: Indiana University, 1973.

Clark, K. B., and Plotkin, L. *The Negro student at integrated colleges.* New York: National Scholarship Service for Negro Students, 1963.

Cleary, T. A. Test bias: Prediction of grades for Negro and white students in integrated colleges. *Journal of Educational Measurement*, 1968, *5*(2), 115–124.

Cleary, T. A., and Hilton, T. L. An investigation of item bias. *Educational and Psychological Measurement*, 1968, *28*(1), 61–75.

Cleary, T. A., Humphreys, L. G., Kendrick, S. A., and Wesman, A. Educational uses of tests with disadvantaged students. *American Psychologist*, 1975, *30*(1), 15–41.

Cole, N. S. Bias in selection. *Journal of Educational Measurement*, 1973, *10*(2), 237–255.

Cronbach, L. J. *Essentials of psychological testing*, 3rd edition. New York: Harper and Row, 1970.

Cronbach, L. J., and Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin*, 1955, *52*(4), 281–302.

Dalton, S. Predictive validity of high school rank and SAT scores for minority students. *Educational and Psychological Measurement*, 1974, *34*(2), 367–370.

Dalton, S. A decline in the predictive validity of SAT and high school achievement. *Educational and Psychological Measurement*, 1976, *36*(2), 445–448.

Darlington, R. B. Another look at "culture fairness." *Journal of Educational Measurement*, 1971, *8*(2), 71–82.

Davis, J. A., and Kerner-Hoeg, S. *Validity of pre-admission indices for blacks and whites in six traditionally white public universities in North Carolina.* Project Report (PR-71-15). Princeton, N.J.: Educational Testing Service, 1971.

Davis, J. A., and Temp, G. Is the SAT biased against black students? *College Board Review*, Fall 1971, No. 81, pp. 4–9.

Dickason, D. G. Predicting the success of freshman engineers. *Personnel and Guidance Journal*, 1969, *47*(10), 1008–1014.

Eagle, N., and Harris, A. S. Interaction of race and test on reading performance scores. *Journal of Educational Measurement*, 1969, *6*(3), 131–135.

Ebel, R. L. *Constructing unbiased achievement tests.* Paper prepared for presentation at the NIE's test bias conference held in Baltimore, Maryland, December 2–4, 1975.

Educational Testing Service. *Report on a study of the use of the National Teacher Examinations by the State of North Carolina.* Princeton, N.J.: Author, August, 1976.

Endler, N. S., and Steinberg, D. Prediction of academic achievement at the university level. *Personnel and Guidance Journal*, 1963, Vol. 41, pp. 694–699.

Equal Employment Opportunity Coordinating Council. Uniform guidelines on employee selection procedures. *Federal Register*, December 30, 1977, *42*(251).

Farver, A. S, Sedlacek, W. E., and Brooks, G. C. Longitudinal predictions of university grades for blacks and whites. *Measurement and Evaluation in Guidance*, 1975, *7*(4), 243–250.

Fincher, C. Is the SAT worth its salt? An evaluation of the use of the Scholastic Aptitude Test in the University System of Georgia over a 13-year period. *Review of Educational Research*, 1974, *44*(3), 293–305.

Flaugher, R. L. The many definitions of test bias. *American Psychologist*, July, 1978, *3*(7), 671–679.

Flaugher, R. L., and Schrader, W. B. *Eliminating differentially difficult items as an approach to test bias.* Research Bulletin (RB-78-0). Princeton, N.J.: Educational Testing Service, March, 1978.

Floyd, W. A. A longitudinal study of the Scholastic Aptitude Test as a predictor of college success. *The School Counselor*, 1967, pp. 138–142.

Ford, S. F., and Campos, S. *Summary of validity data from the Admission Testing Program Validity Study Service.* New York: College Entrance Examination Board, 1977.

Frederiksen, N., and Gilbert, A. C. Replication of a study of differential predictability. *Educational and Psychological Measurement*, 1960, *20*(4), 759–767.

Frederiksen, N., and Melville, J. D. Differential predictability in the use of test scores. *Educational and Psychological Measurement*, 1954, *14*(4), 647–656.

Frederiksen, N., and Schrader, W. B. *A study of 10,000 veteran and nonveteran students in sixteen American colleges.* Princeton, N.J.: Educational Testing Service, 1951.

Fricke, B. G. *Grading, testing, standards, and all that.* Report to the Faculty, Evaluation and Examinations Office, the University of Michigan, 1975, pp. 62–90.

Funches, D. Correlations between secondary school transcript averages and GPA and between ACT scores and GPA of freshmen at Jackson State College. *College and University*, 1967, *43*(1), 52–54.

Goldman, R. D., and Hewitt, B. N. An investigation of test

bias for Mexican-American college students. *Journal of Educational Measurement*, Fall 1975, *12*(3), 187–196.

Goldman, R. D., and Hewitt, B. N. Predicting the success of black, Chicano, Oriental, and white college students. *Journal of Educational Measurement*, 1976, *13*(2), 109–117.

Goldman, R. D., and Richards, R. The SAT prediction of grades for Mexican-American versus Anglo-American students of the University of California, Riverside. *Journal of Educational Measurement*, 1974, *11*(2), 129–135.

Goldman, R. D., and Widawski, M. H. A within subjects technique for comparing college grading standards. Implication in the validity of the evaluation of college achievement. *Educational and Psychological Measurement*, 1976, *36*(2), 381–390.

Green, D. R., and Draper, J. F. *Exploratory studies of bias in achievement tests.* Paper presented to the American Psychological Association, Honolulu, 1972.

Guilford, J. P. *Fundamental statistics in psychology and education: Fourth edition.* New York, N. Y.: McGraw-Hill Inc., 1965.

Gulliksen, H. *When high validity may indicate a faulty criterion.* Paper presented at the annual meetings of the American Psychological Association, Washington, D.C., September 6, 1976. ETS Research Memorandum (RM-76-10). Princeton, N.J.: Educational Testing Service, 1976.

Gulliksen, H., and Wilks, S. S. Regression tests for several samples. *Psychometrika*, 1950, *15*(2), 91–114.

Harris, J., and Reitzel, J. Negro freshmen performance in a predominantly non-Negro university. *Journal of College Student Personnel*, 1967, *8*(6), 366–368.

Hennessy, J. J., and Merrifield, P. R. A comparison of the factor structures of mental abilities in four ethnic groups. *Journal of Educational Psychology*, 1976, *68*(6), 754–759.

Hicks, M. M., Donlon, T. F., and Wallmark, M. M. *Sex differences in item responses on the Graduate Record Examination.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, April, 1976.

Hills, J. R. Prediction of college grades for all public colleges of a state. *Journal of Educational Measurement*, 1964, *1*(2), 155–159.

Hills, J. R., and Stanley, J. C. Easier test improves prediction of black students' college grades. *Journal of Negro Education*, 1970, *39*(4), 320–324.

Hoepfner, R., and Strickland, G. P. *Investigating test bias.* Los Angeles: Center for the Study of Evaluation, University of California, 1972.

Holland, J. L. The prediction of college grades from the California Psychological Inventory and the SAT. *Journal of Educational Psychology*, 1959, *50*(4), 135–142.

Hollmann, T. D. *Differential validity: A problem with tests or criteria?* Paper presented at a meeting of the Midwestern Psychological Association, May, 1973.

Hunter, J. E. *A critical analysis of the use of item means and item-test correlations to determine the presence or absence of content bias in achievement test items.* Paper presented at the National Institute of Education conference on test bias, Annapolis, Md., December 2–5, 1975.

Ironson, G. H. *A comparative analysis of several methods of assessing item bias.* Paper presented at the annual meeting of the American Educational Research Association, Toronto, Can., March, 1978.

Jensen, A. P. *An examination of culture bias in the Wonderlic Personnel Test.* Arlington, Va.: ERIC Clearinghouse, 1973 (ERIC Document Reproduction Service #ED 086-726).

Kallingal, A. The prediction of grades for black and white students at Michigan State University. *Journal of Educational Measurement*, 1971, *8*(4), 263–265.

Lenning, O. T. *Predictive validity of the* ACT *tests at selective colleges.* ACT Research Report, No. 69, August, 1975.

Linn, R. L. Fair test use in selection. *Review of Educational Research*, 1973, *43*(2), 139–161.

Loevinger, J. Objective tests as instruments of psychological theory. *Psychological Reports*, 1957, Vol. 3 (Monograph Supplement N. 3), 635–694.

Lord, F. M. A study of item bias, using characteristic curve theory. In *Basic Problems in Cross-Cultural Psychology.* International Association for Cross-Cultural Psychology, 1976.

Lord, F. M. *Practical applications of item characteristic curve theory.* Research Bulletin (RB-77-3). Princeton, N.J.: Educational Testing Service, 1977.

Mauger, P., and Kolomdin, C. A. Long-term predictive validity of the Scholastic Aptitude Test. *Journal of Educational Psychology*, 1975, *67*(6), 847–851.

McCausland, D. F., and Stewart, N. E. Academic aptitude, study skills and attitudes, and college GPA. *Journal of Educational Research*, 1974, *67*(8), 354–357.

McKelpin, J. A. Some implications of the intellectual characteristics of freshmen entering a liberal arts college. *Journal of Educational Measurement*, 1965, *2*(2), 161–166.

Merz, W. R., and Rudner, L. M. *Bias in testing: A presentation of selected methods.* A paper presented as part of a co-sponsored symposium: Empirical Evaluations of Various Approaches for Identifying Biased Test Item(s), at the annual meeting of the American Educational Research Association and the National Council for Measurement in Education, Toronto, Can., March, 1978.

Messick, S. J. The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 1975, *30*(10), 955–966.

Messick, S. J., and Barrows, T. S. Strategies for research and

evaluation in early childhood education. In *National Society for the Study of Education*, 71st yearbook, Part II, 1972, pp. 261–290.

Michael, W. B., Baker, D., and Jones, R. A. A note concerning the predictive validities of selected cognitive and noncognitive measures for freshmen students in a liberal arts college. *Educational and Psychological Measurement*, 1964, *24*(2), 373–375.

Michael, W. B., and Jones, R. Stability of predictive validities of high school grades and of scores on the Scholastic Aptitude Test of the CEEB for liberal arts students. *Educational and Psychological Measurement*, 1963, *23*(2), 375–378.

Mitchell, J. V. A comparison of the factorial structure of cognitive functions for a high and low status group. *Journal of Educational Psychology*, 1956, *47*(7), 397–414.

Munday, L. Predicting college grades in predominantly negro colleges. *Journal of Educational Measurement*, 1965, *2*(2), 157–160.

Nichols, R. C. Nonintellective predictors of achievement in college. *Educational and Psychological Measurement*, 1966, *26*(4), 899–915.

Nichols, R. C., and Holland, J. L. Prediction of the first year college performance of high aptitude students. *Psychological Monographs*, 1963, *77*(7), 1–29.

Olsen, M. *Summary of main findings on the validity of the* CEEB *tests of developed ability as predictors of college grades*. Statistical Report 57-14. Princeton, N.J.: Educational Testing Service, 1957.

Pandey, R. E. The SCAT and race. *Psychological Reports*, 1971, *28*(2), 459–462.

Passons, W. R. Predictive validities of the ACT, SAT, and high school grades for first semester GPA. *Educational and Psychological Measurement*, 1967, *27*(4), 1143–1144.

Petersen, N. S. *Bias in the selection rule—Bias in the test*. Paper presented at the Third International Symposium on Educational Testing, University of Leyden, The Netherlands, June 28, 1977.

Petersen, N. S., and Novick, M. R. An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, 1976, *13*(1), 3–29.

Peterson, R. E. Predictive validity of a brief test of academic aptitude. *Educational and Psychological Measurement*, 1968, *28*(2), 441–444.

Pfeifer, C. M., and Sedlacek, W. E. The validity of academic prediction for black and white students at a predominantly white university. *Journal of Educational Measurement*, 1971, *8*(4), 253–261.

Potthoff, R. F. *Statistical aspects of the problem of biases in psychological tests*. Institute of Statistics Mimeograph Series No. 479. Chapel Hill: University of North Carolina, Department of Statistics, 1972.

Reck, M. The prediction of achievement in a college science curriculum. *Educational and Psychological Measurement*, 1968, *28*(3), 943–944.

Richards, J. M., Jr., Holland, J. L., and Lutz, S. W. Prediction of student accomplishment in college. *Journal of Educational Psychology*, 1967, *58*, 343–355.

Rock, D. A., and Werts, C. E. *Construct validity of* SAT *across populations: An empirical confirmatory study*. Research Report RR 79-2. Princeton, N.J.: Educational Testing Service, 1979.

Rudner, L. M. *Efforts toward the development of unbiased selection and assessment instruments*. Paper presented at the Third International Symposium on Educational Testing, University of Leyden, The Netherlands, June 28, 1977.

Rudner, L. M., and Convey, J. J. *An evaluation of select approaches for biased item identification*. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Can., March, 1978.

Sassenrath, J. M., and Pugh, R. Relationships among CEEB Scholastic Aptitude Test and American College Test scores and Grade Point Average. *Journal of Educational Measurement*, 1965, *2*(2), 199–205.

Sassenrath, J. M., and Pugh, R. Note: Relationships among CEEB–SAT and ACT scores and GPA: A replication. *Journal of Educational Measurement*, 1966, *3*(1), 37–38.

Scheuneman, J. *A few methods of assessing bias in test items*. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., 1975.

Scheuneman, J. *Validating a procedure for assessing bias in test items in the absence of an outside criterion*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1976.

Schmidt, F. L., and Hunter, J. E. Racial and ethnic bias in psychological tests: Divergent implications of two definitions of test bias. *American Psychologist*, 1974, *29*(1), 1–8.

Schrader, W. B. The predictive validity of College Board Admission Tests. In W. H. Angoff (Ed.), *The College Board Admission Testing Program: A Technical Report on Research and Development Activities Relating to the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board, 1971.

Scott, C. C. *Longer-term predictive validity of college admission tests for Anglo, black and Mexican-American students*. New Mexico: Department of Educational Administration, University of New Mexico, 1976.

Semler, I., and Iscoe, I. Structure of intelligence in Negro and white children. *Journal of Educational Psychology*, 1966, *57*(6), 326–336.

Siegelman, M. SAT and high school average predictions of four year college achievement. *Educational and Psychological Measurement*, 1971, *31*(4), 947–950.

Silverman, B. I., Barton, F., and Lyon, M. Minority group status and bias in college admission criteria. *Educational and Psychological Measurement*, 1976, *36*(2), 401–407.

Stanley, J. C. Predicting college success of the educationally disadvantaged. *Science*, 1971, No. 171, pp. 640–647.

Stanley, J. C., and Hopkins, K. D. *Educational and psychological measurement and evaluation*. Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Stanley, J. C., and Porter, A. C. Correlation of Scholastic Aptitude Test score with college grades for negros versus whites. *Journal of Educational Measurement*, 1967, *4*(4), 199–218.

Stroup, A. L. The prediction of academic performance from personality and aptitude variables. *Journal of Experimental Education*, 1970, *38*(3), 83–86.

Subkoviak, M. J., and Baker, F. B. Test theory. In Shulman, L. S. (Ed.), *Review of Research in Education*. American Educational Research Association, Itasca, Ill.: Peacock Publishers, Inc., 1978, 275–317.

Temp, G. Test bias: Validity of the SAT for blacks and whites in 13 integrated institutions. *Journal of Educational Measurement*, 1971, *8*(4), 245–251.

Thomas, C. L., and Stanley, J. C. Effectiveness of high school grades for predicting college grades of black students: A review and discussion. *Journal of Educational Measurement*, 1969, *6*(4), 203–215.

Thorndike, R. L. Concepts of culture fairness. *Journal of Educational Measurement*, 1971, *8*(2), 63–70.

Veale, J. R., and Foreman, D. I. *Cultural variation in criterion-referenced tests: A "global" item analysis*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, 1976.

Veldman, D. J. Effects of sex, aptitudes and attitudes on the academic achievement of college freshmen. *Journal of Educational Measurement*, 1968, *5*(3), 245–249.

Warren, J. R. *Prediction of college achievement among Mexican-American students in California*. College Board Research and Development Report (RDR-76–77, No. 1; also Research Bulletin (RB-76-22). Princeton, N.J.: Educational Testing Service, 1976.

Wasserman, M. An evaluation of a compensatory introductory sociology section. *Journal of Experimental Education*, 1979, *47*(2).

Wen, S. S., and McCoy, R. E. Relationships of selected non-academic and academic variables to the GPA of black students. *Educational and Psychological Measurement*, 1975, *35*(4), 935–939.

Wild, C. L. Statistical issues raised by Title IX requirements on admission procedures. *Journal of National Association of Women Deans, Administrators, and Counselors*, 1977, Vol. 40, pp. 53–56.

Willingham, W. W. Validity of several methods of expressing high school achievement level. *College and University*, Fall 1964, 49–54.

Willingham, W. W. Predicting success in graduate education. *Science*, 1974, No. 183, 273–278.

Willingham, W. W. *Validity and the Graduate Record Examination Program*. Princeton, N.J.: Educational Testing Service, 1976.

Willis, C. G., and Nicholson, J. Series II SCAT as a college aptitude measure. *Educational and Psychological Measurement*, 1970, *30*(4), 971–975.

Wilson, J. W., and Carry, L. R. Homogeneity of regression—its rationale, computation and use. *American Educational Research Journal*, 1969, *6*(1), 80–90.

Wilson, K. M. *Black students entering CRC colleges: Their characteristics and their first-year academic performance*. Research Memorandum (RM-69-1). Princeton, N.J.: Educational Testing Service, 1969.

Wilson, K. M. *Predicting the long-term performance in college of minority and nonminority students: A comparative analysis in two collegiate settings*. College Board Research and Development Report (RDR-77–78, No. 3); also Research Bulletin (RB-78-6). Princeton, N.J.: Educational Testing Service, April, 1978.

Wood, D. A., and Lebold, W. K. Differential and overall prediction of academic success in engineering: The complementary role of DAT, SAT, and HSR. *Educational and Psychological Measurement*, 1968, *28*(4), 1223–1228.

Zimmerman, W. S. College Entrance Examination Board Scholastic Aptitude Test. In O.K. Buros (Ed.), *The Sixth Mental Measurement Yearbook*. Highland Park, N.J.: Gryphon Press, 1965, pp. 449.

Zimmerman, W. S., and Michael, W. B. A comparison of the criterion-related validities of three college entrance examinations with different content emphases. *Educational and Psychological Measurement*, 1967, *27*(2), 407–412.

Annotated bibliography

Biaggio (1966)

SAT scores for 3,287 entering freshmen in three predominantly black colleges and for 25,674 entering freshmen in 15 predominantly white colleges in Georgia were correlated with their college freshman-year GPAs. Four academic years were covered by the data, 1959–60, 1960–61, 1961–62, and 1962–63. SAT-V median correlations ranged from .38 to .41 for white males, from .46 to .55 for white females, from .38 to .46 for black males, and from .33 to .51 for black females. SAT-M median correlations ranged from .37 to .41 for white males, from .45 to .53 for white females, from .23 to .40 for black males, and from .25 to .43 for black females. No multiple correlation coefficients are reported for SAT-V, SAT-M, and HSGPA as predictors, but median standard errors of estimate are reported.

Borgen (1972)

Black high school seniors who participated in the National Merit program during the 1964–65 year were sent a followup questionnaire in the summer of 1966 asking about their college progress, including their freshman GPA. Out of 4,288 black students who participated, the sample was reduced to 1,314 for whom complete data on both the NMSQT and GPA were available. The undesirable effects of pooling data from a wide variety of college types is demonstrated by computing correlations for several different types of colleges and for all colleges combined. The highest correlation between NMSQT scores and GPA (.54) was obtained for black males attending predominantly black colleges. For the same sample of students, the HSR correlated negatively (−.05) with GPA. When all black males in all colleges were pooled, a very low correlation (.03) between NMSQT score and GPA was obtained. Multiple R's and regression coefficients are also reported.

Bowers (1970)

In September 1968 the University of Illinois at Urbana admitted 515 freshmen, mostly black, as part of a special admissions program. School and College Ability Test (SCAT) scores, high school percentile rank (HSPR), and first-semester GPAs were obtained for both regularly admitted and specially admitted students. The GPAs for the specially admitted students were based on grades earned in regular courses as well as on grades earned in courses tailored to the specially admitted students. The data included 168 men and 237 women specially admitted and 2,938 men and 1,917 women who were regularly admitted. The correlations between SCAT-Verbal and first-semester GPA were .23 for specially admitted men, .25 for regularly admitted men, .17 for specially admitted women, and .28 for regularly admitted women. SCAT-Q correlations with first-semester GPA were .14 for specially admitted men, .31 for regularly admitted men, .11 for specially admitted women, and .21 for regularly admitted women. A similar pattern of predictive relationships was observed when HSPR was used as the predictor: the correlations between HSPR and first-semester GPA were .14 for specially admitted men, .25 for regularly admitted men, .25 for specially admitted women, and .34 for regularly admitted women. Multiple correlation coefficients using SCAT and HSPR as predictors of first-semester GPA were .30 for specially admitted men, .41 for regularly admitted men, .29 for specially admitted women, and .39 for regularly admitted women.

Breland (1977)

Group comparisons for male, female, majority, and minority students were conducted for the Test of Standard Written English (TSWE). Data for two academic years and from 18 different institutions were analyzed by pooling data across institutions within each of the academic years. Analyses of data from the first academic year focused on relationships between TSWE scores and grades. Analyses of the second academic year focused on relationships between TSWE scores and scores on graded writing samples. No important group differences in traditional correlational analyses for either grade or essay prediction were observed. No important group differences were observed in either correlational analyses or regression analyses of the second year essay data. Analyses of correct and incorrect placement decisions (hits and misses) at specific TSWE cut-off scores revealed no noteworthy group differences whether outcomes were based on English course grades or on freshman year writing performance. Whether grades or essays were used as the outcome, the proportion of incorrect decisions (misses) was less for minority students than for any of the groups. For all groups, the TSWE appeared to predict freshman year writing performance as well as or better than pre-course writing samples, high school English grades, or high school rank in class.

Calkins and Whitworth (1974)

Data were collected for 3,237 students who entered the University of Texas at El Paso between the fall of 1969 and the fall of 1971. These students consisted of 711 males and 530 females with Spanish surnames, and of 1,192 males and 804 females with non-Spanish surnames. SAT-V scores, SAT-M scores, quartile rank in high school class, and freshman GPA were obtained and related. Correlations of SAT-V and college GPA were .25 for Spanish-surname males, .33 for Spanish-surname females, .32 for non-Spanish-surname males, and .37 for non-Spanish-surname females. Correlations of SAT-M and freshman GPA were .23 for Spanish-surname males, .28 for Spanish-surname-females, .34 for non-Spanish-surname males, and .34 for non-Spanish-surname females. Correlations of an SAT composite and freshman GPA were .26 for Spanish-surname males, .35 for Spanish-surname females, .38 for non-Spanish-surname males, and .40 for non-Spanish-surname females. Analysis of regression coefficients indicated no statistically significant differences across groups with the exception of males vs. females. There was some indication that females performed better in college than would be predicted from a common regression equation combining both sexes.

Cleary (1968)

Data were collected and analyzed for blacks and whites in three integrated colleges. In Schools 1 and 2 FGPA was used as the criterion. In school 3 the latest cumulative GPA was used as the criterion. Predictors used were the SAT-V, SAT-M, and high school rank-in-class (HSR), when available. Correlations between predictors and criteria were computed and analyses of covariance conducted to compare regression lines for blacks and whites with respect to slopes and intercepts. The results for School 1 showed similar correlations between SAT-V and GPA for three groups (blacks, whites matched with blacks on curriculum, and a random white group), but SAT-M had its lowest correlation in the black group (.01) and its highest in the random white group (.41). In School 2 none of the correlations were high with HSR for blacks being an especially poor predictor of GPA (.02). In School 3, where the cumulative GPA was used as the criterion, correlations were much higher with HSR appearing to be the best predictor for all groups (r's in excess of .60). SAT-V and SAT-M also correlated well with GPA for all groups. The analyses of regression lines indicated significant differences in one of the three schools. When black means were used in regression equations developed for either white or combined groups, the predicted GPA for blacks was in every case higher than the actual attained black GPA.

Cleary, Humphreys, Kendrick, and Wesman (1975)

This paper reviews the issue of test use and differences in performance for different ethnic groups. Organized in five sections, the paper presents a comprehensive definition of abilities, summarizes some common types of test misuse and misinterpretation, notes the kinds of statistical information needed to use tests effectively, and includes a discussion of alternatives to ability tests. The first section gives the background from which the paper developed. It is observed that test misuse is easier to correct than test misinterpreta-

tion. Different kinds of test validation are described and related to the issue of test fairness. Tables from Cleary (1968) of intercorrelations for SAT-V, SAT-M, HSR, and college GPA and tables describing regression equations are reprinted. Alternatives to tests discussed include lotteries, use of prior experience, and demographic categories. But it is suggested that fairness may be more of a problem with alternatives than with tests themselves. The use of diagnostic tests and the development of normative data for disadvantaged groups are emphasized, and the study of moderator variables are seen as promising approaches to the societal dilemma considered in the paper.

Dalton (1974)

The relative value of high school achievement and test scores as predictors of college grades was investigated. Subjects were 87 males and 114 females admitted as freshmen at Indiana University in the fall of 1972 under a special program that waives normal entrance requirements. Instead, these students were admitted by recommendations of high school counselors, social workers, and school officials who were in contact with disadvantaged high school students, especially intercity blacks. The selected group consisted of 169 blacks, 17 Mexican Americans or Puerto Ricans, and 15 whites. High school rank (HSR) correlated .20 with GPA, .02 with SAT-V scores, .07 with SAT-Q scores, .06 with SAT-total scores. None of the low SAT correlations were statistically significant, however, and, therefore, could have resulted from chance variations. A secondary analysis of multiple correlations suggested that SAT and HSR combined as predictors of college GPA work best for students with low HSR and high SAT scores (R = .62) than for students with high HSR and high SAT scores (R = .36), for students with low HSR and high SAT scores (R = .35), or for students with low HSR and low SAT-scores (R = .29). The problem of restriction in the range of abilities resulting from such subgroupings was not discussed.

Davis and Kerner-Hoeg (1971)

Blacks enrolled in six traditionally white universities in North Carolina during the 1966–70 period were studied. Intercorrelations among SAT scores, high school rank-in-class, and term grades alone were computed separately for black vs. white samples. Additionally, comparisons were made of regression lines for the two groups. The number of black students available for the study ranged from a low of 32 students in one of the six institutions to a high of 97 in another. Whites were randomly sampled to obtain (nominally) 100 males and 100 females in each institution. For the white samples (males and females combined), multiple correlations of SAT-V, SAT-M, and HSR with first term college grades ranged from a low of .41 to a high of .56. For the black samples the multiple R's ranged from .31 to .70. Regression analyses showed that the use of regression equations developed from the white samples would predict GPAs for blacks that were higher than those actually achieved by blacks in all six of the institutions.

Davis and Temp (1971)

This paper combines data and analyses from 13 institutions studied by Temp (1971) and six institutions studied by Davis and Kerner-Hoeg (1971) to allow for comparisons of blacks and whites in a total of 19 institutions. Multiple correlations of SAT-V and SAT-M as predictors of freshman grades are reported. Those multiple correlations that were statistically significant ranged from .23 to .55 for whites and from .25 to .65 for blacks. A more critical test of group differences was provided by the comparison of regression lines. The regression comparisons suggested that different prediction equations would be appropriate for blacks and whites in most of the 19 institutions, because the white equation or an equation derived from combined groups would tend to overpredict college performance for blacks.

Farver, Sedlacek, and Brooks (1975)

This was a study of all new freshman black students who entered the University of Maryland in the fall of 1968 (n = 126) and in the fall of 1969 (n = 133). Samples of white students were drawn for comparison. Multiple correlations and equations were determined for the prediction of freshman, sophomore, junior, and senior year GPA, both individually and cumulatively. The predictors used were SAT-V, SAT-M, and high school GPA. Most of the predictors correlated significantly with college GPA for most groups and most criteria. High school GPA was an especially poor predictor for black males. SAT-V was positively correlated with college grades for

all groups, was the most consistent predictor over the years for all groups, and tended to be best for black males. Additionally, it was suggested that the academic predictors used in the study had considerable usefulness in predicting grades beyond the freshman year. Some of the 1968 data and analyses reported here were previously reported in Pfeifer and Sedlacek (1971).

Funches (1967)

The relative predictive capability of the ACT composite score and high school GPA (HSGPA) was studied for 540 black freshmen enrolled at Jackson State College in the fall of 1965. The correlation obtained between the ACT composite score and first-term college GPA was .36. The correlation obtained between HSGPA and first-term college GPA was .06. The comparison of ACT scores and HSGPA was also made by computing the mean first-term GPAs for groups identified by four ACT score groupings:

N	ACT score range	ACT mean	HSGPA mean	College GPA
87	1–6	5.10	2.31	.86
295	7–12	9.18	2.15	1.03
136	13–18	14.82	2.75	1.17
22	19–24	20.20	3.17	1.84

The HSGPA mean did not conform to the pattern of abilities indicated by either the ACT means or the college GPA means. The 87 students scoring in the lowest ACT score group and having the lowest college GPA mean did not have the lowest observed HSGPA mean. It was suggested that HSGPA is not acceptable as a predictor of college performance at Jackson State College.

It was concluded that the ACT composite score is a more reliable predictor of first term success at Jackson State College than HSGPA.

Goldman and Hewitt (1975)

Chicanos and Anglos were compared in four California state colleges. In School 1, there were 261 Chicano and 5,635 Anglo cases with adequate data; in School 2, 84 Chicano and 5,500 Anglo cases; in School 3, 180 Chicano and 2,926 Anglo cases; in School 4, 131 Chicano and 3,127 Anglo cases. Predictors were HSGPA, SAT-V, and SAT-M, and the criterion was GPA. Correlations are given for the predictors acting both individually and in combination. It was observed that the HSGPA was the best individual predictor of GPA, generally, that SAT-V and SAT-M tended to predict better for Anglos, and that the intercorrelations among predictors was markedly higher for Chicanos. An analysis of regression systems for the groups suggested that they were essentially the same for all groups. No systematic over- or under-predictions was observed when Chicano means were used in Anglo regression equations.

Goldman and Hewitt (1976)

Analyses were conducted of HSGPA, SAT-V, SAT-M, GPA, and major field selections for 272 black, 852 Oriental, 188 Chicano, and 4,259 white undergraduates enrolled during the 1973–74 academic year at UCLA. Correlation coefficients with college GPA obtained by group were:

Group	N	HSGPA	SAT-V	SAT-M	R
Black	272	.27	.28	.18	.33
Oriental	852	.35	.18	.35	.42
Chicano	188	.37	.16	.17	.38
White	4,259	.36	.30	.21	.43

Another interesting correlational comparison across groups was that between SAT-V scores and SAT-M scores. The SAT-V/SAT-M correlation for blacks (.68) and Chicanos (.62) were somewhat higher than those for Orientals (.39) and whites (.44). Observed SAT-V and SAT-M standard deviations for blacks (110,108) and Chicanos (99,101) were slightly larger than those for Orientals (98,92) and whites (87,92). Predictions using white-derived regression equations were accurate for the other groups:

Group	Predicted GPA	Actual GPA
Black	2.52	2.52
Oriental	2.80	2.81
Chicano	2.61	2.65

It was reported that similar findings were obtained in three other California colleges, but complete results were not given.

Goldman and Richards (1974)

Two studies are reported comparing Chicanos and Anglos at the University of California at Riverside. The first included all students who entered the university in the fall of 1971 (n = 210). The second study was an attempted replication of the first study with a much larger sample (n = 1700). Second quarter GPA was used as the criterion and SAT-V and SAT-M as the predictors. The results of the first study showed similar zero-order correlations of SAT-V and GPA, but differences for SAT-M (.37 vs. .12, with the lower r for Chicanos). No significant differences were obtained in the multiple correlation using both SAT-V and SAT-M as predictors; however, differences between Chicano and Anglo regression systems were indicated. Substantial overpredictions of Chicano GPA occurred when the Anglo regression equation was used. The results of the second study were similar to those of the first except that the multiple correlation was higher for the Anglo group (.33 vs. .24, $p < .05$). As in the first study, the Anglo regression equation would overpredict Chicano GPA. Additionally, the low SAT-M weights observed for Chicanos in the first study were replicated.

Goldman and Widawski (1976)

Data on black, white, and Chicano undergraduates enrolled during the spring of 1974 at UCLA, UC Davis, UC San Diego, and UC Irvine were analyzed. Relative effects of using the SAT and HSGPA singly and in combination were investigated both through multiple regression analyses and through analyses of selection-outcome matrices (four-fold tables generated by dichotomizing at hypothetical success-fail and accept-reject points). Analyses indicated that little addition to explained college GPA variance of Chicanos was contributed by the use of SAT scores in addition to HSGPA, but that the use of SAT scores measurably improved the prediction of black college GPA in the colleges studied. The selection-outcome matrix analyses suggested that little improvement in outcomes (over that obtainable with HSGPA alone) occurred as a result of the use of the SAT. The selection-outcome analyses also indicated that the use of the SAT increased the proportion of Chicanos and blacks classified as false negatives (rejected, but would have succeeded).

Harris and Reitzel (1967)

Prediction equations for 45 black freshmen enrolled at an integrated, predominantly white university during the years 1961 through 1964 were compared with equations based on the total of 3,895 freshman students for these years. Three types of prediction equations were developed: a prediction based on high school rank (HSR) alone, a prediction based on SAT-V and SAT-M scores combined, and a prediction based on the combination of HSR, SAT-V, and SAT-M. The black and total equations based on HSR alone resulted in a mean black predicted GPA of 1.63 and 2.06, respectively. SAT-V and SAT-M combined resulted in a mean predicted GPA for blacks of 2.12 and of 2.07, respectively. Using all three predictors, mean black predicted GPAs of 1.90 (blacks equation) and 2.04 (total equation) were obtained. It was concluded that prediction equations based on the total population overpredicted black college performance, and that HSR was the principal source of the overprediction. Correlations were also reported: between HSR and freshman GPA for blacks (.50), between HSR and freshman GPA for the total of 3,895 students (.44).

Hills (1964)

Data from all publicly supported colleges in Georgia for the five years from 1958 through 1962 were analyzed. The data were from 19 colleges and for 27,961 students. Three of the colleges were traditionally black colleges. The measures investigated were: SAT-V, SAT-M, HSGPA, and first-year college GPA. The average multiple correlation obtained for all colleges over all five years was .71 for females and .60 for males. For all students combined, the average correlation between HSGPA and college GPA was .55. Addition of SAT-V and SAT-M to the multiple prediction increased the average R to .65. For the three traditionally black colleges, the r's and R's were as shown at the top of next page.

College	1958 r HSGPA	1958 R	1959 r HSGPA	1959 R	1960 r HSGPA	1960 R	1961 r HSGPA	1961 R	1962 r HSGPA	1962 R
B (males)	.34	.38	.43	.49	.32	.59	.52	.56	.50	.61
(females)	.58	.64	.56	.69	.31	.38	.60	.72	.50	.60
F (males)	.33	.50	.38	.57	.52	.58	.52	.66	.63	.69
(females)	.62	.71	.60	.70	.66	.69	.55	.67	.57	.65
M (males)	.28	.51	.41	.59	.42	.50	.54	.66	.51	.59
(females)	.61	.76	.55	.62	.53	.56	.63	.70	.39	.44

These correlations suggested that the addition of test scores to HSGPA consistently improved prediction of college performance in traditionally black colleges. The author notes that restriction in range on SAT scores in colleges B, F, and M did not appreciably affect the multiple correlations. The average standard deviation for these colleges was approximately 45 on SAT-V and 48 on SAT-M vs. 106 and 111, respectively, for all colleges combined).

Hills and Stanley (1970)

To test an hypothesis that an easier test would result in improved prediction of college performance for blacks, the three traditionally black colleges of the University System of Georgia (which all required the SAT) participated in a special study in 1966 in which Level 4 (for grades 6–8) of the School and College Ability Test (SCAT) was also administered during the orientation program of the freshman year. The SCAT score distributions obtained were nearly symmetrical with means at the 60th percentile of 8th grade norms for SCAT-V and at the 35th percentile for SCAT-Q. The results of multiple correlational analyses predicting college GPA were:

College	Sex	N	R's SAT-V + SAT-M	SCAT-V + SCAT-Q	SAT-V + SAT-M + HSGPA	SCAT-V SCAT-Q HSGPA
1	Male	101	.22	.43	.50	.56
	Female	215	.47	.60	.63	.68
2	Male	154	.39	.54	.65	.70
	Female	213	.44	.53	.61	.63
3	Male	116	.36	.53	.51	.61
	Female	198	.54	.64	.65	.70

These results were reported to indicate clearly that the SCAT (Level 4) was a better predictor of first-year college grades in traditionally black colleges than the SAT. The authors note that the improved prediction occurred despite the shorter length of the SCAT (60 items for the SCAT-V vs. 90 items for the SAT-V; 50 items for the SCAT-Q vs. 60 items for the SAT-Q) suggesting that a longer SCAT would provide an even better prediction.

Kallingal (1971)

A two-year GPA criterion was used to compare prediction of college performance for 225 black and 511 randomly selected white students who entered Michigan State University in the fall of 1968 and completed the winter term in 1970. The predictors were five tests: English and Reading tests developed by the Office of Evaluation Services at Michigan State; Verbal, Informational, and Numerical subtests of the College Qualification Test (CQT). Observed standard deviations for the black sample were larger for the English, Reading, Verbal, and Informational tests but smaller for the Numerical test. The observed standard deviation for the GPA criterion was slightly larger for the white sample (.52 vs. .47). Correlations for the predictive relationship between tests and GPA were:

Test	r's Blacks	r's Whites	R's Blacks	R's Whites
MSU English	.42	.43		
MSU Reading	.47	.47		
CQT Verbal	.43	.37	.51	.53
CQT Informational	.38	.38		
CQT Numerical	.20	.31		

As observed in some other studies, the observed intercorrelations among the test scores were generally higher for the black sample than for the white sample. A comparison of black and white regressions showed that the white regression equation significantly overestimated black criterion performance in agreement with Cleary's (1968) results for a Southwest college and Temp's (1971) results in 12 of 13 integrated colleges.

Linn (1973)

This is a frequently cited review of the various definitions of test bias and proposed psychometric models for fair selection. It also includes some reanalyses of the data of Cleary (1968), Davis and Kerner-Hoeg (1971), and Temp (1971), applying the selection model suggested by Thorndike (1971). The definitions of fairness reviewed are those of Cleary (1968), Thorndike (1971), Cole (1973), and Darlington (1971). The data reanalyses show that while a common regression equation overpredicts black college performance, the use of such equations in isolation from other factors can result in a smaller proportion of black students admitted than could be successful. Thus, the data reanalyses confirmed the Thorndike contention. Linn observes that there will always be a conflict between the Thorndike and Cleary definitions of fairness as long as there exist differences in the mean performance of the two groups being compared. Linn argues that it is necessary to consider all the implications of admissions practices rather than to consider predicted performance only.

McKelpin (1965)

Data were collected for students entering North Carolina College at Durham (primarily black) in 1961, 1962, and 1963. First semester average grades were used as the criterion and SAT-V, SAT-M, and HSGPA were studied as predictors. These results are reported:

	Predictors				
Year	SAT-V	SAT-M	HSGPA	SAT-V SAT-M HSGPA	N
Males:					
1961	.52	.44	.43	.60	40
1962	.49	.47	.53	.64	182
1963	.59	.41	.39	.66	100
Females:					
1961	.52	.32	.57	.66	69
1962	.47	.55	.46	.64	339
1963	.48	.52	.49	.67	100

It is concluded that the predictive validities for the predictors used are as high in this primarily black college as those usually reported for college freshmen generally.

Munday (1965)

Five traditionally black southern colleges were studied with respect to the value of ACT scores and HSGPA as predictors of college GPA. The results were then compared to results from national data collections conducted by the American College Testing Program for the year (1964) in which data were collected. The results are summarized in the following table:

		Correlations with GPA of		
College	N	ACT Composite	HSGPA	ACT and HSGPA
1	105	.62	.46	.66
2	205	.54	.50	.60
3	233	.58	.51	.65
4	561	.56	.39	.58
5	554	.48	.38	.52
National	103,256 (204 colleges)	.52 (median r)	.56 (median r)	.63 (median R)

Standard deviations for the ACT composite score for the five black colleges were 4.24, 3.78, 4.46, 4.03, and 4.02, respectively, and are slightly less than the standard deviation for the national sample (4.77). Standard devia-

tions of high school grades in four areas were substantially below that for the national sample, while standard deviations of college GPA were substantially below the national for some curriculum areas (English, natural science) but not for some of the others (mathematics, social science).

Pandey (1971)

A total of 47 black and 33 white graduating college seniors were studied by correlating SCAT percentile ranks with cumulative college GPA. Race, sex, and sex within race comparisons were made, even though the number of cases available was limited. The correlation for 37 females was .49 but that for 43 males only .06 and not statistically significant. Nonsignificant correlations of −.32, .27, .12, and .66 were obtained for black males, black females, white males, and white females, respectively. This study is somewhat limited by the small sample sizes.

Peterson (1968)

The purpose of this paper was to report relationships between a newly developed test, the Control Test for Academic Aptitude (CTAA), and freshman grades. Of more interest for the present review, however, are correlations reported between SAT-V scores, SAT-M scores, and freshman GPA for 225 blacks in a Georgia liberal arts college for men and for 438 whites in two northern liberal arts colleges. Correlations with student-reported high school GPA and rank are also reported for each of the three colleges. SAT-V and SAT-M scores did not correlate highly with the self-reported high school performance for blacks, but they did correlate with freshman GPA for blacks. The observed predictive correlations were highest in a Minnesota Roman Catholic college for women, second highest were in the Georgia college for black men, and the lowest in a Wisconsin coeducational college. The observed concurrent correlations (self-reported high school performance vs. SAT performance) were also highest for the students in the Minnesota Roman Catholic college but lowest for the black southern college.

Pfeifer and Sedlacek (1971)

Subjects were freshmen entering the University of Maryland at College Park in September, 1968. All blacks entering in that year who were identifiable and who had complete data were included in the study (n = 126). A total of 178 white students were randomly sampled from the entering class. With freshman GPA as the criterion, analyses were made of SAT-V, SAT-M, and HSGPA as predictors. Results indicated that high school grades did not correlate as highly with college grades for black males as for the other three groups studied. In the multiple regression equation the SAT-V was the predictor of primary importance for black males rather than high school grades. Separate equations for race-sex groupings gave the most accurate predictions of college performance with blacks being overpredicted by equations based on a random sample of students. Further reporting on this same sample, but using longer-range criteria, is provided in Farver, Sedlacek, and Brooks (1975).

Scott (1976)

Anglo (N = 878), black (N = 67), and Mexican American (N = 66) samples were compared with respect to data collected on junior-year college GPA, high school GPA, and ACT composite scores. Correlational analyses yielded the following predictive relationships:

Predictor	Anglo	Black	Chicano
HSGPA	.31	.36	.26
ACT	.21	−.04	.20
HSGPA and ACT	.32	.38	.29

It was concluded that, when ACT scores and HSGPA are used in combination, long-term predictive validity is achieved for all groups. The author notes, however, considerable attrition in the original sample of 1,926 Anglos, 154 blacks, and 264 Chicanos. Analyses of covariance using the Gulliksen-Wilks (1950) procedure revealed few important differences in errors of estimate, slopes, or intercepts for the three groups when both HSGPA and ACT scores were used as predictors.

Silverman, Barton, and Lyon (1976)

Both the Thorndike and Cleary definitions of bias in college admissions measures were examined for 135 black, 272 white, and 121 Jewish students at Roosevelt University. Regression equations were developed separately for each group and for the total group for predicting both first year and second year GPA from ACT

scores and high school percentile rank (HSR). Multiple correlations of .31, .44, and .37 were obtained for blacks, whites, and Jews, respectively, for the first year GPA criterion and of .49, .39, and .49, respectively, for the second year GPA criterion. Regression equations developed for the total sample significantly overpredicted black performance and significantly underpredicted Jewish performance for both years. Using the Thorndike definition of bias and considering "surpassing C" as the criterion of success, the ratios of predicted to actual numbers of students surpassing C for the first year were 1.44, 1.18, .78, respectively. For the second year the ratios were 1.88, 1.30, 1.17, respectively. Like the regression equations, these ratios suggest that Jews, but not blacks, are discriminated against in admissions. However, the use of separate regression equations would have resulted in selection biased against blacks according to the Thorndike model.

Stanley and Porter (1967)

This study extended the study of Biaggio (1966) from four years to six years. SAT-V and SAT-M scores were obtained for students at 14 predominantly white colleges and three predominantly black colleges in Georgia. Correlations between both SAT-V scores and freshman grades and SAT-M scores were found to be generally higher for white women than for white men, black women, or black men. SAT-V tended to correlate slightly better with grades than SAT-M for all groups. Multiple correlations combining SAT-V, SAT-M, and high school grades as predictors of freshman GPA showed white women to be best predicted for all of the six years studied (1959–60 through 1964–65). Average multiple correlation coefficients for the six years were .72 for white women, .60 for white men, .63 for black women, and .60 for black men. Although Biaggio had made corrections for restriction of range, it was argued that such corrections are inappropriate when the restrictions are caused primarily by the difficulty of the test.

Temp (1971)

Blacks and whites enrolled as full-time freshmen in thirteen integrated institutions during the 1968–69 academic year were compared. SAT-V and SAT-M scores were computed separately and in combination as predictors of freshman-year GPA. The sample of white students was drawn from the same class and curriculum area from which the black student sample was drawn. A Gulliksen-Wilks test for equality of regression planes was conducted for all thirteen institutions. The results strongly suggested that different regression systems were appropriate for blacks and whites in at least 10 of the institutions. The primary differences in the regressions were in the intercepts, and usually the intercept was lower for blacks than for whites. When the regression equations developed from the white sample were used to predict for the black sample, overpredictions occurred in 12 of the 13 institutions. That is, use of the white equation falsely indicated that blacks would perform better than they actually did perform.

Thomas and Stanley (1969)

This paper reviews evidence and presents additional data concerning the relative predictive value of high school grades and test scores for black students. The authors cite studies by McKelpin (1965), Munday (1965), Funches (1967), Cleary (1968), and Peterson (1968) as supportive of the notion that the high school record is less useful as a predictor of college performance for black students than admissions test scores. Additional analyses are given for 736 black students at a predominantly black university showing that the high school record (HSR) was the poorest predictor of college grades for three different criteria (overall FGPA, English GPA, and math GPA). For 536 black students who took math, the HSR correlated only .14 with math GPA while the SAT-M correlated .40. Both SAT-V and SCAT-V correlated .43 with English GPA while HSR correlated only .22. The results reviewed and presented indicate that high school grades do not consistently make the greatest contribution (as they do for whites) in predicting college grades of black students, particularly male black students.

Warren (1976)

Mexican American students at three universities of the California State University and Colleges System were compared with other students by developing regression equations for the prediction of college grades from test scores and high school grades for each group. Both freshman GPA and longer-range criteria were used. And both SAT and ACT scores were used as predictors. With

rare exceptions, regression equations for Mexican Americans did not differ from those for other students. It was concluded, therefore, that the use of separate regression equations for Mexican American applicants was not supported by the results of the study.

Wasserman (1978)

Using a large section of an introductory sociology course, educationally disadvantaged students (black and Puerto Rican) were compared with regular students with respect to the predictive relationship between SAT scores and relative sociology class standing in the form of Z-scores. For the disadvantaged group of 61 students the SAT-V score correlated .65 and the SAT-M score .50 with sociology class standing. For 706 regular students, the SAT-V correlated .64 and the SAT-M .44 with sociology class standing. High school rank correlated only .27 with sociology class standing for disadvantaged students and only .35 for regular students. The author suggests that the high predictive correlations for the SAT-V may have been the result of several factors: the concentration of the study within a course and discipline requiring verbal skills in particular, the use of the class standing criterion in the form of Z-scores that maximized the criterion variance, and the difficulty of the reading materials used in the course.

Wilson (1969)

This was a study of black women students who entered several highly selective women's liberal arts colleges in 1965, 1966, and 1967. During these three years a total of 168 black women entered the selective institutions studied. Data collected included responses to a 200-item questionnaire, SAT-V scores, SAT-M scores, scores on College Board Achievement Tests, high school class rank, and college freshman GPA. Analyses compare the black students to all students entering in 1966. The SAT scores, the Achievement Test scores, and the high school rank were combined to predict college freshman GPA for black students and for all students entering in 1966. The predicted college mean GPA was compared with the college GPA actually attained for both groups in four different colleges. In three of these four colleges the observed mean predicted GPA for blacks was higher than the observed mean actual GPA (no statistical tests reported). But when all students were compared with respect to predicted and actual performance, the observed mean predicted GPA was less than the observed mean actual GPA (no statistical tests reported). Thus, it appeared that the black women's performance tended to be overpredicted while students, generally, tended to be underpredicted by the regression equation.

Wilson (1978)

Data from two institutions were collected and analyzed with the objective of comparing minority and nonminority student performance over both the short-range (first year) and the long-range (four-year cumulative) in relation to performance on several predictor variables (SAT-V, SAT-M, high school rank, College Board Achievement Tests average). At the first institution, a highly selective Eastern college, data for a total of 121 minority students (self-identified blacks, American Indians, Latins, or Oriental Americans) were compared with a random sample of 1,133 nonminority students. At the second institution, the main campus of a Midwestern state university system, data for a total of 272 minority students were compared with a 20 percent random sample of 1,003 nonminority students. Multiple and zero-order correlations are reported for both samples between predictors and cumulative as well as individual semester GPAs. Comparisons of regression systems for the minority and nonminority samples are also reported. Predictive correlations differed for the two settings, with semester 8 GPA at the selective institution correlating higher with SAT-V for nonminority students than for minority students (.29 vs. .09); at the nonselective institution, in contrast, semester 8 GPA correlated less with SAT-V for nonminority students than for minority students (.32 vs. .47). Four-year cumulative GPA correlated higher with SAT-V in both groups at the selective institution (.45 vs. .23) and at the nonselective institution (.42 vs. .56). Multiple R's using all predictors and cumulative GPA were .57 (nonminority) and .42 (minority) at the selective institution and .57 (nonminority) and .69 (minority) at the nonselective institution.

Table A-1. Correlations between high school grades, high school ranks, and test scores (as predictors) and college GPA (as criterion) for general (unspecified) populations

Study	Number of cases	Year	Test	r's: Test score M	F	T	HSR M	F	T	R's: HSGPA M	F	T	HSGPA and test M	F	T	HSR and test M	F	T
Astin (1971)																		
College 1	138F	–	SAT		.43						.39			.52				
			NMSQT		.30									.45				
College 2	172M, 118F	–	SAT	.36	.33					.50	.48		.55	.54				
			NMSQT	.30	.33								.53	.54				
College 3	160M	–	SAT	.55						.54			.62					
			NMSQT	.51									.61					
College 4	267M, 260F	–	SAT	.30	.29					.45	.51		.48	.52				
			NMSQT	.35	.37								.50	.52				
College 5	571M, 513F	–	SAT	.33	.27					.49	.43		.52	.46				
			NMSQT	.29	.27								.51	.45				
College 6	444M, 245F	–	SAT	.40	.37					.41	.37		.51	.50				
			NMSQT	.41	.42								.52	.53				
College 7	191M	–	SAT	.31						.49			.54					
			NMSQT	.26									.53					
College 8	398M	–	SAT	.38						.61			.64					
			NMSQT	.38									.64					
College 9	365F	–	SAT		.36						.38			.47				
			NMSQT		.31									.44				
College 10	249M	–	SAT	.37						.41			.49					
			NMSQT	.34									.48					
College 11	207F	–	SAT		.35						.64			.66				
			NMSQT		.39									.66				
College 12	227M, 164F	–	SAT	.30	.32					.33	.49		.45	.53				
			NMSQT	.25	.33								.41	.52				
College 13	739M, 250F	–	SAT	.34	.36					.47	.36		.54	.48				
			NMSQT	.32	.34								.53	.44				
College 14	190M, 190F	–	SAT	.32	.37					.45	.51		.49	.55				
			NMSQT	.25	.35								.46	.55				
College 15	238M, 200F	–	SAT	.40	.46					.46	.38		.55	.53				
			NMSQT	.36	.36								.52	.46				
College 16	544M, 228F	–	SAT	.40	.37					.51	.53		.59	.57				
			NMSQT	.32	.35								.55	.56				
College 17	185M	–	SAT	.51						.63			.70					
			NMSQT	.42									.65					
College 18	194M, 185F	–	SAT	.50	.58					.74	.66		.75	.72				
			NMSQT	.48	.61								.75	.71				
College 19	211M, 150F	–	SAT	.51	.61					.48	.67		.59	.73				
			NMSQT	.46	.57								.58	.72				
			ACT	.57	.73								.46	.57				
College 20	91M, 170F	–	ACT	.58	.43					.59	.51		.68	.57				
			NMSQT	.52	.47								.65	.59				
College 21	382M, 433F	–	ACT	.50	.49					.59	.59		.63	.63				
			NMSQT	.44	.44								.62	.62				

Table A-1 continued

Study	Number of cases	Year	Test	r's									R's					
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Brown and Lightsey (1970)																		
Georgia Southern College	76M, 125F	1969	SAT-V	.54	.50													
			SAT-M	.46	.45													
			SAT-T	.56	.55													
Chansky (1965)																		
State U. of North Carolina (Raleigh)																		
School 1	41	1963	SAT-V			.20			.33									
			SAT-M			.17												
School 2	68		SAT-V			.24			.03									
			SAT-M			.38												
School 3	42		SAT-V			.38			.26									
			SAT-M			.47												
School 4	124		SAT-V			.16			.45									.49
		–	SAT-M			.26												
School 5	83		SAT-V			.51			.59									.66
		–	SAT-M			.35												
School 6	64		SAT-V			.38			.50									.65
		–	SAT-M			.29												
			SAT															.58
Chase and Thompson (1973)																		
Midwestern college	2,247M, 2,284F	–	SAT-V	.38	.42		.45	.46										
			SAT-M	.39	.45													
Dalton (1976)																		
Indiana U.	740M, 740F	1961	SAT													.62	.64	
	1,202M, 1,340F	1965														.52	.56	
	2,063M, 2,054F	1970														.53	.58	
	2,247M, 2,284F	1973														.51	.54	
	2,374M, 2,489F	1974														.47	.49	
Dickason (1969)																		
Cornell U. (Engr.)	596	1967	SAT-V			.27												
			SAT-M			.32												
	369		NMSQT			.39												
Endler and Steinberg (1963)																		
Liberal arts college in Ontario	45M, 24F	1962	SCAT-V	.35	.55					.49	.84							
			SCAT-Q	.02	.49													
			SCAT-T	.27	.59													

Table A-1 continued

Study	Number of cases	Year	Test	r's: Test score M	F	T	HSR M	F	T	R's: HSGPA M	F	T	HSGPA and test M	F	T	HSR and test M	F	T
Fincher (1974)																		
U. System of Ga.	Unspecified	1958–66,	SAT-V	.36	.50					.49	.60		.59	.70				
(19 colleges)		1970	SAT-M	.36	.47													
			SAT	.41	.56													
Floyd (1967)																		
Southern women's college	239F	1960	SAT		.63													
Ford and Campos[a] (1977)	57–62	1964	SAT-V	.33	.44	.37	(see note c)			.52	.57	.55	.56	.65	.62	(see note c)		
	colleges		SAT-M	.32	.42	.32												
(Median correlations)	61–80	1965	SAT-V	.32	.41	.36				.48	.57	.56	.54	.63	.61			
	colleges		SAT-M	.31	.38	.29												
	68–90	1966	SAT-V	.29	.38	.33				.47	.55	.53	.52	.61	.57			
	colleges		SAT-M	.28	.34	.28												
	62–101	1967	SAT-V	.30	.40	.36				.45	.55	.52	.51	.62	.57			
	colleges		SAT-M	.28	.36	.28												
	41–49	1968	SAT-V	.30	.44	.40				.48	.54	.56	.57	.63	.62			
	colleges		SAT-M	.29	.40	.32												
	45–73	1969	SAT-V	.28	.38	.39				.46	.56	.53	.52	.61	.59			
	colleges		SAT-M	.30	.35	.33												
	129–154	1970	SAT-V	.29	.39	.37				.43	.52	.49	.49	.58	.56			
	colleges		SAT-M	.27	.36	.29												
	32–51	1971	SAT-V	.30	.37	.34				.44	.47	.48	.50	.55	.53			
	colleges		SAT-M	.26	.39	.28												
	78–101	1972	SAT-V	.30	.40	.35				.45	.50	.50	.52	.59	.57			
	colleges		SAT-M	.32	.41	.33												
	58–89	1973	SAT-V	.35	.40	.38				.44	.47	.50	.54	.55	.56			
	colleges		SAT-M	.36	.40	.36												
	70–110	1974	SAT-V	.35	.40	.42				.44	.47	.50	.52	.57	.58			
	colleges		SAT-M	.37	.42	.39												
Frederiksen and Gilbert (1960)																		
Princeton U. (Engr.)	107	1958	SAT-V			.37			.41									
			SAT-M			.27												
			CEEB-English			.31												
			CEEB-Science			.42												
Frederiksen and Melville (1954)																		
Princeton U. (Engr.)	154	—	SAT-V			.29			.44									
			SAT-M			.37												
			CEEB-Science			.37												

Table A-1 continued

Study	Number of cases	Year	Test	r's						R's								
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Frederiksen and Schrader (1951)																		
Eastern college																		
Veterans	531	1946–47	SAT-V			.41												
			SAT-M			.33												
			SAT			.46												
Nonveterans	694		SAT-V			.39												
			SAT-M			.30												
			SAT			.46												
Harris and Reitzel (1967)																		
Anonymous North Carolina college	3,985	1961–64							.44									
Hills (1964)[b]																		
U. System of Ga. (19 colleges)	27,961																	
		1958	SAT							.47	.60	.54	.61	.72	.67			
		1959	SAT							.51	.61	.56	.61	.71	.67			
		1960	SAT							.47	.61	.55	.58	.70	.64			
		1961	SAT							.54	.60	.57	.61	.72	.67			
		1962	SAT							.51	.58	.54	.61	.68	.64			
Holland (1959)																		
National samples		1957																
standard	476M, 185F		SAT-V	.12	.26													
			SAT-M	.17	.15													
Cross-validation	481M, 179F		SAT-V	.04	.06													
			SAT-M	.15	.00													
Lenning (1975)																		
Air Force Academy	271	1967–68	SAT-V			.28												
			SAT-M			.29												
			SAT			.35												
			ACT			.44												
	348	1968–69	SAT-V			.20												
			SAT-M			.22												
			SAT			.26												
			ACT			.41												
Mauger and Kolomdin (1975)																		
Unspecified liberal arts college	318	–	SAT-V			.50												
			SAT-M			.46												
Seniors (Cumulative GPA)	520		SAT-V			.26												
			SAT-M			.22												

Table A-1 continued

Study	Number of cases	Year	Test	r's						R's								
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
McCausland and Stewart (1974)																		
U. of Wyoming	47M, 59F	–	ACT			.66						.71			.79			
Michael, Baker, and Jones (1964)																		
U. of Southern California	210M, 177F	1958	SAT-V	.35	.44					.44	.52							
			SAT-M	.24	.39													
			SAT	.35	.48													
Michael and Jones (1963)																		
U. of Southern California	186M, 234F	1956	SAT-V	.45	.53					.58	.47		.60	.60				
			SAT-M	.46	.34													
			SAT	.46	.53													
	310M, 272F	1957	SAT-V	.29	.45					.40	.51		.47	.59				
			SAT-M	.22	.39													
			SAT	.31	.47													
	209M, 177F	1958	SAT-V	.35	.41					.44	.52		.44	.59				
			SAT-M	.24	.39													
			SAT	.33	.48													
	209M, 233F	1960	SAT-V	.25	.41					.40	.52		.44	.61				
			SAT-M	.23	.20													
			SAT	.28	.41													
	228M, 288F	1961	SAT-V	.27	.29					.35	.48		.42	.53				
			SAT-M	.16	.16													
			SAT	.27	.29													
Munday (1965)[a]																		
National norm data	103,256	–	ACT			.52						.56			.63			
Nichols (1966)																		
National sample																		
Merit Finalists	179M, 138F	1961	NMSQT	.10	–.03		.14	.08										
			SAT-V	.17	.01													
			SAT-M	.10	–.10													
Normative sample	201M, 218F		NMSQT	.18	.23		.31	.07										
Nichols and Holland (1963)																		
National Merit Finalists	544M, 275F	1960–61					.24	.21										

Table A-1 continued

Study	Number of cases	Year	Test	r's									R's					
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Olsen (1957)																		
College A	656M	1956	TDA	.45									.64					
			SAT	.42									.62					
College B	501M		TDA	.56									.65					
			SAT	.51									.64					
College C	265M		TDA	.34									.47					
			SAT	.25									.45					
College D	150M		TDA	.61									.67					
			SAT	.57									.70					
College E	138M		TDA	.46									.61					
			SAT	.49									.60					
College F	63M		TDA	.55									.63					
			SAT	.21									.42					
College G	308F		TDA		.45									.50				
			SAT		.36									.45				
College H	157F	—	TDA		.29									.32				
			SAT		.18									.23				
College I	155F	—	TDA		.42									.51				
			SAT		.36									.48				
College J	135F	—	TDA		.51									.66				
			SAT		.54									.63				
College K	130F	—	TDA		.54									.62				
			SAT		.62									.66				
College L	113F	—	TDA		.37									.63				
			SAT		.40									.58				
College M	472	—	TDA			.37									.43			
			SAT			.32									.39			
College N	197	—	TDA			.52									.66			
			SAT			.49									.62			
College O	180	—	TDA			.42									.53			
			SAT			.39									.51			
College P	75	—	TDA			.26									.38			
			SAT			.26									.45			
Passons (1967)																		
Fresno State College	882	1963–64	SAT-V			.39						.41						
			SAT-M			.26												
			SAT			.19												
			ACT			.36												
Peterson (1968)																		
Wisconsin	106	1965	Control test-AA			.39			.46			.49						
			SAT-V			.29												
			SAT-M			.24												

Table A-1 continued

Study	Number of cases	Year	Test	r's						R's								
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Reck (1968)																		
Purdue U.	109M, 71F	—	SAT-V													.67	.59	
(Science)			SAT-M													.72	.62	
Richards, Holland, and Lutz (1967)																		
American college survey	503M, 592F	1962–63, 1964																
28 institutions																		
Freshmen	503M, 592F		ACT-soc. stu.	.29	.44					.32	.37							
Sophomores	1,373M, 1,419F		ACT-soc. stu.	.44	.50					.39	.43							
			ACT-English		.48													
Sassenrath and Pugh (1965)																		
Indiana U.	80	1963–64	SAT-V			.59												
			SAT-M			.35												
			SAT			.53												
			ACT-English			.59												
			ACT-math			.27												
			ACT-soc. stu.			.51												
			ACT-na. sci.			.39												
			ACT			.53												
Sassenrath and Pugh (1966)																		
Indiana U.	177	1964–65	SAT-V			.50												
			SAT			.49												
			ACT-English			.55												
			ACT			.53												
Schrader (1971)[a]	116M, 143F		SAT-V	.33	.41	.39												
			SAT-M	.30	.36	.33	.47	.54	.55							.55	.62	.62
Siegelman (1971)																		
City College of New York	80M, 95F																	
Freshmen			SAT-V	−.01	.27					.25	.31		.21	.41				
			SAT-M	−.05	.29													
Sophomores			SAT-V	.20	.38					.25	.30		.20	.46				
			SAT-M	−.09	.31													
Juniors			SAT-V	.01	.29					.39	.36		.35	.48				
			SAT-M	.09	.33													
Seniors			SAT-V	.05	.24					.25	.33		.29	.43				
			SAT-M	.16	.33													

Table A-1 continued

Study	Number of cases	Year	Test	r's: Test score M	F	T	HSR M	F	T	HSGPA M	F	T	R's: HSGPA and test M	F	T	HSR and test M	F	T
Silverman, Barton, and Lyon (1976)																		
Roosevelt U.	528	1967–70	ACT															
Freshmen																		.40
Sophomores																		.35
Stroup (1970)																		
College of Wooster	968M, 970F	1959–63	SAT-V	.47	.43													
		1959–63	SAT-M	.43	.36													
			SAT	.52	.47													
Thomas and Stanley (1969)	332F		SAT-V		.51			.56										
			SAT-M		.43													
	106		SAT-V			.29			.46									
			SAT-M			.24												
Veldman (1968)																		
U. of Texas	2,315	—	SAT-V			.43												
			SAT-M			.41												
			SAT			.49												
Willingham (1964)																		
Georgia Tech.	761	1957	SAT						.48			.47			.60			.61
	880	1960	SAT						.47			.48			.62			.61
Willis and Nicholson (1970)																		
U. of Missouri	112	1968	SAT-V			.26												
			SAT-M			.28												
			SCAT-V			.33												
			SCAT-Q			.32												
			SCAT-total			.38												
Wilson (1978)																		
Anon. College ("College")																		
Freshman GPA	1,248	1970–71	SAT-V			.40			.34									.51
			SAT-M			.40												
			CB Ach. avg.			.50												
Soph cum GPA	1,158	1970–72	SAT-V			.45			.35									
			SAT-M			.39												
			CB Ach. avg.			.53												
Jnr cum GPA	1,013	1970–73	SAT-V			.46			.35									
			SAT-M			.39												
			CB Ach. avg.			.53												
Snr cum GPA	810	1970–74	SAT-V			.50			.34									
			SAT-M			.40												
			CB Ach. avg.			.34												

87

Table A-1 continued

Study	Number of cases	Year	Test	r's: Test score M	F	T	HSR M	F	T	R's: HSGPA M	F	T	HSGPA and test M	F	T	HSR and test M	F	T
Sem. 7 GPA	1,055	1973–74	SAT-V			.35			.19									
			SAT-M			.27												
			CB Ach. avg.			.38												
Sem. 8 GPA	808	1973–74	SAT-V			.32			.22									
			SAT-M			.19												
			CB Ach. avg.			.34												
Wilson (1978)																		
Anon. University ("Main")																		
Sem. 1 GPA	1,266	1971–72	SAT-V			.40			.51									
			SAT-M			.41												
Freshman GPA	1,173	1971–72	SAT-V			.44			.55									.57
			SAT-M			.44												
Total cum GPA	1,275	1971–75	SAT-V			.48			.56									.57
			SAT-M			.45												
Wood and Lebold (1968)																		
Purdue U. (Engr.)	616	1966	SAT-V			.25			.33									.48
			SAT-M			.38												
			DAT-CSA			.05												
			DAT-SR			.12												
			DAT-MR			.16												
			DAT-AR			.18												
Zimmerman[a] (in Buros, 1965)		1959–61	SAT-V	.35	.36													
			SAT-M	.33	.26													
Zimmerman and Michael (1967)																		
State college in California (LA)	410M, 413F	1963	SAT							.18			.28	.50				
			ACT							.18	.46		.33	.54				

a. Median correlations only reported.
b. Average correlations only reported.
c. HSR and HSGPA were both used. Correlations are recorded under HSGPA, however.

88

Table A-2. Correlations between high school grades, high school ranks, and test scores (as predictors) and college performance (as criterion) for white male and female samples

Study	Number of cases	Year	Test	r's			R's								
				Test score			HSR			HSGPA			HSGPA nd test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Biaggio (1966)[a]																		
Public colleges in Georgia	25,674	1959–60	SAT-V	.39	.55													
			SAT-M	.41	.53													
		1960–61	SAT-V	.41	.54													
			SAT-M	.37	.49													
		1961–62	SAT-V	.34	.51													
			SAT-M	.38	.51													
		1962–63	SAT-V	.38	.46													
			SAT-M	.37	.45													
Bowers (1970)																		
University of Illinois	2,938M, 1,917F	1968	SCAT-V	.25	.28													
			SCAT-Q	.31	.21		.35	.34								.41	.39	
Calkins and Whitworth (1974)																		
U. of Texas	1,996	1969–71	SAT-V	.32	.37		.42	.38								.47	.46	
			SAT-M	.33	.34													
			SAT	.38	.40													
Cleary (1968)																		
School 1	118		SAT-V			.45												
			SAT-M			.41												
School 2	365		SAT-V			.38			.38									
			SAT-M			.30												
School 3	2,181		SAT-V			.47			.66			.64						
			SAT-M			.39												
Davis and Kerner-Hoeg (1971)																		
Six colleges in North Carolina																		
F	100M, 101F		SAT-V			.19			.37									.41
			SAT-M			.21												
G	100M, 100F		SAT-V			.35			.43									.53
			SAT-M			.24												
H	92M, 83F		SAT-V			.24			.55									.55
			SAT-M			.15												
I	96M, 97F		SAT-V			.31			.51									.54
			SAT-M			.28												
J	100M, 100F		SAT-V			.39			.42									.56
			SAT-M			.37												
K	86M, 101F		SAT-V			.29			.33									.43
			SAT-M			.20												

Table A-2 continued

Study	Number of cases	Year	Test	r's									R's					
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Davis and Temp (1971)																		
19 unspecified colleges																		
College 1	100	1969	SAT			.27												
College 2	100		SAT			.23												
College 3	100		SAT			.38												
College 4	100		SAT			.33												
College 5	140		SAT			.55												
College 6	100		SAT			.39												
College 7	100		SAT			.15												
College 8	100		SAT			.51												
College 9	100		SAT			.45												
College 10	100		SAT			.25												
College 11	69		SAT			.43												
College 12	100		SAT			.49												
College 13	100		SAT			.46												
College 14	100		SAT			.25												
College 15	200		SAT			.38												
College 16	175	1969	SAT			.27												
College 17	200		SAT			.34												
College 18	200		SAT			.45												
College 19	189		SAT			.31												
Farver, Sedlacek, and Brooks (1975)																		
		1968																
U. of Maryland		*Entrants*																
Freshman GPA	79M, 99F	1968–69	SAT										.63[c]	.65[c]				
Soph GPA	55M, 79F	1969–70	SAT										.61	.73				
Soph cum GPA	55M, 79F	1968–70	SAT										.68	.76				
Junior GPA	44M, 66F	1970–71	SAT										.45	.50				
Jnr cum GPA	44M, 66F	1968–71	SAT										.57	.73				
Senior GPA	34M, 56F	1971–72	SAT										.28	.39				
Snr cum GPA	34M, 56F	1968–72	SAT										.55	.67				
		1969																
		Entrants																
Freshman GPA	70M, 52F	1969–70	SAT										.61	.65				
Soph GPA	56M, 66F	1969–71	SAT										.50	.42				
Soph cum GPA	56M, 66F	1969–71	SAT										.59	.49				
Junior GPA	45M, 36F	1969–72	SAT										.53	.36				
Jnr cum GPA	45M, 36F	1969–72	SAT										.64	.50				

90

Table A-2 continued

Study	Number of cases	Year	Test	r's			R's								
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Goldman and Hewitt (1976)																		
UCLA	4,259	1973–74	SAT-V			.30						.36			.43			
			SAT-M			.21												
Four Calif. state colleges		1974–75																
School 1	5,635		SAT-V			.16						.37			.43			
			SAT-M			.13												
School 2	5,500		SAT-V			.42						.42			.48			
			SAT-M			.38												
School 3	2,926		SAT-V			.28						.43			.47			
			SAT-M			.24												
School 4	3,127		SAT-V			.22						.36			.41			
			SAT-M			.18												
Goldman and Richards (1974)																		
UC Riverside	210	1971	SAT-V			.40												
			SAT-M			.37												
			SAT			.44												
	1,700	1973	SAT-V			.30												
			SAT-M			.25												
			SAT			.33												
Harris and Reitzel (1967)																		
U. of North Carolina	3,895	1961–64							.44									
Kallingal (1971)																		
Michigan State U. (MSU)	511	1968–70	CQT-V			.37												
			CQT-Info.			.38												
			CQT-Num.			.31												
			CQT-comp.			.53												
			MSU-English			.43												
			MSU-Reading			.47												
Peterson (1968)	332F		Control															
College B		1965	test-AA	.51			.56			.59								
			SAT-V	.51														
			SAT-M	.43														
Pfeifer and Sedlacek (1971)																		
U. of Maryland	79M, 99F	1968	SAT-V	.51	.39					.58	.65		.63[d]	.65[d]				
			SAT-M	.41	.35													

91

Table A-2 continued

Study	Number of cases	Year	Test	r's									R's					
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Scott (1976)																		
Southwestern University																		
Junior GPA	878		ACT			.21						.31			.32			
Silverman, Barton, and Lyon (1976)																		
Roosevelt University	272	1967–70																
Freshman GPA																		.44
Sophomore GPA																		.39
Stanley and Porter[b] (1967)																		
14 colleges in Georgia		1959–60	SAT-V	.42	.54								.62	.72				
			SAT-M	.43	.50													
		1960–61	SAT-V	.40	.55								.58	.72				
			SAT-M	.37	.50													
		1961–62	SAT-V	.35	.50								.60	.71				
			SAT-M	.36	.49													
		1962–63	SAT-V	.38	.47								.59	.70				
			SAT-M	.38	.45													
		1963–64	SAT-V	.31	.50								.59	.72				
			SAT-M	.31	.49													
		1964–65	SAT-V	.34	.48								.57	.70				
			SAT-M	.42	.42													
Warren (1976)																		
Three state colleges in California																		
College A	255																	
Freshman GPA			ACT										.13	.49	.32			
Sophomore GPA			ACT										.56	.35	.49			
Junior GPA			ACT												.31			
Senior GPA			ACT												.33			
College B	514																	
Freshman GPA			ACT										.29	.49	.41			
Freshman GPA			SAT										.43	.28	.32			
College C	1,575																	
Upper div. GPA			SAT												.49			
Junior GPA			ACT												.32			
Senior GPA			ACT												.34			
Wasserman (1978)																		
New Jersey college	774		SAT-V			.64			.35									
			SAT-M			.44												

Table A-2 continued

Study	Number of cases	Year	Test	r's: Test score M	F	T	HSR M	F	T	HSGPA M	F	T	R's: HSGPA and test M	F	T	HSR and test M	F	T
Wilson (1978)																		
Anonymous college ("College")																		
Freshman GPA	1,103	1970–71	SAT-V			.34			.32									.51
			SAT-M			.30												
			CB Ach. avg.			.45												
Soph cum GPA	1,048	1970–72	SAT-V			.38			.34									
			SAT-M			.32												
			CB Ach. avg.			.48												
Jnr cum GPA	916	1970–73	SAT-V			.41			.33									
			SAT-M			.32												
			CB Ach. avg.			.48												
Snr cum GPA	738	1970–74	SAT-V			.45			.32									
			SAT-M			.32												
			CB Ach. avg.			.49												
Sem. 6 GPA	911	1972–73	SAT-V			.29			.22									
			SAT-M			.19												
			CB Ach. avg.			.32												
Sem. 7 GPA	953	1973–74	SAT-V			.32			.18									
			SAT-M			.22												
			CB Ach. avg.			.34												
Sem. 8 GPA	724	1973–74	SAT-V			.29			.20									
			SAT-M			.13												
			CB Ach. avg.			.30												
Anonymous university ("Main")																		
Sem. 1 GPA	968	1971–72	SAT-V			.39			.50									
			SAT-M			.39												
Freshman GPA	922	1971–72	SAT-V			.41			.53									.57
			SAT-M			.40												
Total cum GPA	968	1971–75	SAT-V			.42			.53									.57
			SAT-M			.40												

a. Median correlations only reported.
b. Correlations computed from Fisher's Z values reported.
c. Also reported in Pfeifer and Sedlacek (1971).
d. Also reported in Farver, Sedlacek, and Brooks (1975).

Table A-3. Correlations between high school grades, high school ranks, and test scores (as predictors) and college performance (as criterion) for black (or predominantly black) male and female samples

Study	Number of cases	Year	Test	r's: Test score M	F	T	HSR M	F	T	HSGPA M	F	T	R's: HSGPA and test M	F	T	HSR and test M	F	T
Biaggio (1966)[a]																		
Three colleges in Georgia	3,287	1959–60	SAT-V	.46	.43													
			SAT-M	.40	.40													
		1960–61	SAT-V	.38	.41													
			SAT-M	.23	.25													
		1961–62	SAT-V	.40	.51													
			SAT-M	.46	.43													
		1962–63	SAT-V	.43	.33													
			SAT-M	.39	.37													
Bowers (1970)																		
U. of Illinois	168M, 237F	1968	SCAT-V	.23	.17		.14	.25								.30	.29	
			SCAT-Q	.14	.11													
Cleary (1968)																		
School 1	59		SAT-V			.47												
			SAT-M			.01												
School 2	83		SAT-V			.26			.02									
			SAT-M			.17												
School 3	125		SAT-V			.47			.61			.51						
			SAT-M			.47												
Dalton (1974)																		
Indiana U.	201	1972	SAT-V			.02			.20									
			SAT-M			.07												
Davis and Kerner-Hoeg (1971)																		
Six colleges in N. Carolina																		
F	97	1966–69	SAT-V			.13			.22									.31
			SAT-M			.26												
G	67		SAT-V			.06			.46									.50
			SAT-M			.30												
H	32		SAT-V			.26			.22									.36
			SAT-M			.15												
I	41		SAT-V			.19			.45									.46
			SAT-M			.09												
J	43		SAT-V			.43			.06									.49
			SAT-M			.39												
K	38		SAT-V			.59			.40									.70
			SAT-M			.40												

Table A-3 continued

Study	Number of cases	Year	Test	r's						R's								
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	r	F	T	M	F	T	M	F	T	M	F	T
Davis and Temp (1971)																		
19 unspecified colleges																		
College 1	100		SAT			.26												
College 2	100		SAT			.15												
College 3	100		SAT			.30												
College 4	100		SAT			.18												
College 5	140		SAT			.51												
College 6	100		SAT			.25												
College 7	100		SAT			.07												
College 8	100		SAT			.27												
College 9	100		SAT			.06												
College 10	100		SAT			.39												
College 11	69		SAT			.08												
College 12	39		SAT			.41												
College 13	100		SAT			.15												
College 14	100		SAT			.26												
College 15	67		SAT			.31												
College 16	34		SAT			.28												
College 17	41		SAT			.19												
College 18	43		SAT			.48												
College 19	30		SAT			.65												
Farver, Sedlacek, and Brooks (1975)		*1968 Entrants*																
Freshman GPA	64M, 62F	1968–69	SAT-V and M										.61[b]	.67[b]				
Sophomore GPA	54M, 62F	1969–70											.39	.63				
Soph cum GPA	54M, 62F	1968–70											.51	.71				
Jnr GPA	43M, 37F	1970–71											.47	.36				
Jnr cum GPA	43M, 37F	1968–71											.52	.61				
Snr GPA	32M, 27F	1971–72											.40	.35				
Snr cum GPA	32M, 27F	1971–72											.52	.42				
		1969 Entrants																
Freshman GPA	58M, 75F	1969–70	SAT-V and M										.44	.57				
Sophomore GPA	48M, 59F	1970–71											.31	.50				
Soph cum GPA	48M, 59F	1969–71											.47	.59				
Jnr GPA	36M, 40F	1971–72											.23	.43				
Jnr cum GPA	36M, 40F	1969–72											.78	.61				
Funches (1967)																		
Jackson State	540	1965	ACT			.36				.06								
Goldman and Hewitt (1976)																		
UCLA	272	1973–74	SAT-V			.28				.27					.33			
			SAT-M			.18												

Table A-3 continued

Study	Number of cases	Year	Test	r's: Test score M	F	T	HSR M	F	T	HSGPA M	F	T	R's: HSGPA and test M	F	T	HSR and test M	F	T
Goldman and Widawski (1976)																		
UCLA	353	1974	SAT									.28			.32			
UC Davis	192											.46			.54			
UC San Diego	196											.26			.32			
UC Irvine	132											.14			.32			
Harris and Reitzel (1967)																		
U. of N. Carolina	45	1961–64							.50									
Hills (1964)																		
Three colleges in Georgia																		
College B		1958	SAT-V and M							.34	.58		.38	.64				
		1959	SAT-V and M							.43	.56		.49	.69				
		1960	SAT-V and M							.32	.31		.59	.38				
		1961	SAT-V and M							.52	.60		.56	.72				
		1962	SAT-V and M							.50	.50		.61	.60				
College F		1958	SAT-V and M							.33	.62		.50	.71				
		1959	SAT-V and M							.38	.60		.57	.70				
		1960	SAT-V and M							.52	.66		.58	.69				
		1961	SAT-V and M							.52	.55		.66	.67				
		1962	SAT-V and M							.63	.57		.69	.65				
College M		1958	SAT-V and M							.28	.61		.51	.71				
		1959	SAT-V and M							.41	.55		.59	.62				
		1960	SAT-V and M							.42	.58		.50	.56				
		1961	SAT-V and M							.54	.63		.66	.70				
		1962	SAT-V and M							.51	.39		.59	.44				
Hills and Stanley (1970)																		
Three colleges in Georgia		1966–67																
College 1	101M, 215F		SAT	.22	.47								.50	.63				
			SCAT	.43	.60								.56	.68				
College 2	154M, 213F		SAT	.39	.44								.65	.61				
			SCAT	.54	.53								.70	.63				
College 3	116M, 198F		SAT	.36	.54								.51	.65				
			SCAT	.53	.64								.61	.70				
Kallingal (1971)																		
Michigan State U.	225	1968–70	CQT-Verbal			.43												
			CQT-Info.			.38												
			CQT-Num.			.20												
			CQT-comp.			.51												
			MSU-English			.42												
			MSU-Reading			.47												

Table A-3 continued

Study	Number of cases	Year	Test	r's						R's								
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
McKelpin (1965)																		
North Carolina College—Durham	40M, 69F	1961	SAT-V	.52	.52					.43	.57		.60	.66				
			SAT-M	.44	.32													
	182M, 339F	1962	SAT-V	.49	.47					.53	.46		.64	.64				
			SAT-M	.47	.55													
	100M, 100F	1963	SAT-V	.59	.48					.39	.49		.66	.67				
			SAT-M	.41	.52													
Munday (1965)																		
School 1	105	1964	ACT			.62						.46			.66			
School 2	205	1964	ACT			.54						.50			.60			
School 3	233	1964	ACT			.58						.51			.65			
School 4	561	1964	ACT			.56						.39			.58			
School 5	554	1964	ACT			.48						.38			.52			
Pandey (1971)																		
Snr cum GPA	16		SCAT	−.32	.27													
Peterson (1968)																		
College in Georgia	225M	1965	Control test-AA	.40			.41			.18								
			SAT-V	.30														
			SAT-M	.35														
Pfeifer and Sedlacek (1971)																		
U. of Maryland	64M, 62F	1968	SAT-V	.56	.47													
			SAT-M	.46	.49					.23	.60		.61[c]	.67[c]				
Scott (1976)																		
Southwestern U. Junior GPA	67		ACT			−.04						.36			.36			
Silverman, Barton, and Lyon (1976)																		
Roosevelt U.	135	1967–70																
Freshmen			ACT															.31
Sophomores			ACT															.49
Stanley and Porter[d] (1967)																		
Three colleges in Georgia		1959–60	SAT-V	.38	.43								.54	.67				
			SAT-M	.34	.43													
		1960–61	SAT-V	.41	.32								.55	.55				
			SAT-M	.25	.26													
		1961–62	SAT-V	.42	.50								.63	.69				
			SAT-M	.40	.41													

97

Table A-3 continued

Study	Number of cases	Year	Test	r's									R's					
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Stanley and Porter (1967)																		
Three colleges in Georgia		1962–63	SAT-V	.41	.32								.63	.57				
			SAT-M	.41	.36													
		1963–64	SAT-V	.41	.41								.61	.68				
			SAT-M	.33	.43													
		1964–65	SAT-V	.53	.40								.65	.59				
			SAT-M	.40	.36													
Thomas and Stanley (1969)																		
Unspecified college	736	1965	SAT-V			.35			.29									.48
			SAT-M			.31												
			SCAT-V			.35												.48
			SCAT-Q			.32												
			COOP-R			.33												
			COOP-E			.37												
			COOP-M			.35												
Wasserman (1978)																		
College in New Jersey	61	1973–74	SAT-V			.65			.27									
			SAT-M			.50												
Wen and McCoy (1975)																		
Jackson State	164M, 202F		ACT	.48	.50													
Wilson (1978)																		
Anon. college ("College")																		
Freshman GPA	116	1970–71	SAT-V			.24			.24									.42
			SAT-M			.19												
			CB Ach avg.			.36												
Soph cum GPA	110	1970–72	SAT-V			.25			.19									
			SAT-M			.14												
			CB Ach avg.			.32												
Jnr cum GPA	97	1970–73	SAT-V			.18			.26									
			SAT-M			.13												
			CB Ach avg.			.32												
Snr cum GPA	72	1970–74	SAT-V			.24			.15									
			SAT-M			.18												
			CB Ach avg.			.32												
Sem. 7 GPA	102	1973–74	SAT-V			.11			.07									
			SAT-M			.03												
			CB Ach avg.			.24												
Sem. 8 GPA	72	1973–74	SAT-V			.09			.16									
			SAT-M			.01												
			CB Ach avg.			.13												

Table A-3 continued

Study	Number of cases	Year	Test	r's: Test score M	F	T	HSR M	F	T	HSGPA M	F	T	R's: HSGPA and test M	F	T	HSR and test M	F	T
Wilson (1978)																		
Anon. University ("Main")																		
Sem. 1 GPA	270	1971–72	SAT-V			.37			.50									
			SAT-M			.44												
Freshman GPA	251	1971–72	SAT-V			.44			.56									.60
			SAT-M			.44												
Total cum GPA	275	1971–75	SAT-V			.56			.59									.69
			SAT-M			.55												

a. Median correlations only reported.
b. Reported also by Pfeifer and Sedlacek (1971).
c. Also reported by Farver, Sedlacek, and Brooks (1975).
d. Correlations computed from Fisher's Z values reported.

Table A-4. Correlations between high school grades, high school ranks, and test scores (as predictors) and college performance (as criterion) for Mexican American or Chicano male and female samples

Study	Number of cases	Year	Test	r's: Test score M	F	T	HSR M	F	T	HSGPA M	F	T	R's: HSGPA and test M	F	T	HSR and test M	F	T
Calkins and Whitworth (1974)																		
U. of Texas	1,241	1969–71	SAT-V	.25	.33		.36	.34								.38	.41	
			SAT-M	.23	.28													
			SAT	.26	.35													
Goldman and Hewitt (1975)																		
Four California state colleges		1974–75																
School 1	261		SAT-V			.16						.37			.38			
			SAT-M			.13												
School 2	84		SAT-V			.42						.42			.49			
			SAT-M			.38												
School 3	180		SAT-V			.25						.21			.28			
			SAT-M			.14												
School 4	131		SAT-V			.22						.36			.37			
			SAT-M			.18												
Goldman and Hewitt (1976)																		
UCLA	188	1973–74	SAT-V			.16						.37			.38			
			SAT-M			.17												

Table A-4 continued

Study	Number of cases	Year	Test	r's									R's					
				Test score			HSR			HSGPA			HSGPA and test			HSR and test		
				M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
Goldman and Richards (1974)																		
UC Riverside	42	1971	SAT-V			.33												
			SAT-M			.12												
	110	1973	SAT-V			.29												
			SAT-M			.10												
			SAT			.24												
Goldman and Widawski[a] (1976)																		
UCLA	263	1974	SAT									.37			.37			
UC Davis	84	1974	SAT									.41			.49			
UC San Diego	180	1974	SAT									.20			.24			
UC Irvine	129	1974	SAT									.36			.36			
Scott (1975)																		
Southwestern U.																		
Junior GPA	66		ACT			.20						.26			.29			
Warren (1976)																		
Three state colleges in California																		
College A Fresh GPA	262	1967–70	ACT										.58	.56	.55			
College A Soph GPA			ACT										.58	.46	.48			
College A Jnr GPA			ACT												.32			
College A Snr GPA			ACT												.11			
College B Fresh GPA	232	1967–70	ACT										.35	.51	.43			
College B Fresh GPA			SAT										.13	.37	.31			
College C Up. Div GPA	168	1967–70	SAT												.41			
College C Jnr GPA			ACT												.29			
College C Snr GPA			ACT												.52			

a. Correlations computed from R^2 reported in author's table 1.

Table A-5. List of abbreviations

Tests	Other terms
ACE — American Council on Education (aptitude)	GPA — grade point average (college)
ACT — American College Testing Program (aptitude): soc. stu. — social studies na. sci. — natural sciences comp. — composite	HSGPA — high school grade point average
	HSR — high school rank
CAT — Cooperative Achievement Tests (achievement)	M — males
CQT — College Qualification Tests (aptitude): info. — information num. — numerical	F — females
	T — combined male and female samples, or total
DAT — Differential Aptitude Test (aptitude): AR — abstract reasoning CSA — clerical speed and accuracy MR — mechanical reasoning SR — spatial relations	
CB — College Entrance Examination Board (achievement): (English, science)	
NMSQT — National Merit Scholastic Qualification Test (aptitude)	
PSAT — Preliminary Scholastic Aptitude Test (aptitude)	
SAT — Scholastic Aptitude Test (aptitude): V — verbal M — mathematics comp. — composite	
SCAT — School and College Aptitude Test (aptitude): V — verbal Q — quantitative T — total	
STEP — Sequential Tests of Educational Progress (achievement)	
TDA — Tests of Developed Abilities (achievement)	